Intuitive experimentation in the physical world

Citation for published version:
Bramley, NR, Gerstenberg, T, Tenenbaum, JB & Gureckis, TM 2018, 'Intuitive experimentation in the physical world', *Cognitive Psychology*, vol. 105, pp. 9-38. https://doi.org/10.1016/j.cogpsych.2018.05.001

Digital Object Identifier (DOI):
10.1016/j.cogpsych.2018.05.001

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Cognitive Psychology

Intuitive Experimentation in the Physical World

Neil R. Bramley[1], Tobias Gerstenberg[2],
Joshua B. Tenenbaum[2], and Todd M. Gureckis[1]
[1]Department of Psychology, NYU, New York, NY {neil.bramley, todd.gureckis}@nyu.edu;
[2]Center for Brains, Minds and Machines, MIT, Cambridge, MA {tger, jtb}@mit.edu

Author Note

The authors thank to Jack Valenti and Victor Wang for help with video coding and David Lagnado, Anselm Rothe and Tomer Ullman for useful comments, and Hongyi Zhang for initial code. Preliminary analysis of Experiment 1 appeared in a non-archival conference paper Bramley, Gerstenberg, and Tenenbaum (2016).

NB is supported by a Moore Sloan Data Science Environment postdoc position at NYU as well as a James S. McDonnell Scholar Award to TMG. TG and JT are supported by the Center for Brains, Minds & Machines (CBMM), funded by NSF STC award CCF-1231216 and by an ONR grant N00014-13-1-0333. TMG is supported by BCS-1255538 from the National Science Foundation and the John S. McDonnell Foundation Scholar Award.

Abstract

Many aspects of our physical environment are hidden. For example, it is hard to estimate how heavy an object is from visual observation alone. In this paper we examine how people actively "experiment" within the physical world to discover such latent properties. In the first part of the paper, we develop a novel framework for the quantitative analysis of the information produced by physical interactions. We then describe two experiments that present participants with moving objects in "microworlds" that operate according to continuous spatiotemporal dynamics similar to everyday physics (i.e., forces of gravity, friction, etc...). Participants were asked to interact with objects in the microworlds in order to identify their masses, or the forces of attraction/repulsion that governed their movement. Using our modeling framework, we find that learners who freely interacted with the physical system selectively produced evidence that revealed the physical property consistent with their inquiry goal. As a result, their inferences were more accurate than for passive observers and, in some contexts, for yoked participants who watched video replays of an active learner's interactions. We characterize active learners' actions into a range of micro-experiment strategies and discuss how these might be learned or generalized from past experience. The technical contribution of this work is the development of a novel analytic framework and methodology for the study of interactively learning about the physical world. Its empirical contribution is the demonstration of sophisticated goal directed human active learning in a naturalistic context.

Keywords: active learning; mental simulation; experimental design; physical understanding

Intuitive Experimentation in the Physical World

What makes physics physics is that experiment is intimately connected to theory. It's one whole.

— LENE HAU

Much of what we believe about the world, we infer from passive observation and inductive reasoning. For example, if we see that the ground is wet, we might infer that it has been raining. However, we also continuously shape our experience by actively interacting with the world. To determine if a container holds water or sand, we might shake it and observe the resulting forces and sounds. From a causal learning perspective, our actions can be seen as *interventions* (Pearl, 2000) that help reveal how the world works. As such, our everyday actions can share some of the characteristics of scientific experiments: comparing the outcomes of different manipulations while controlling for confounding factors. For example, we might lift two suitcases to judge which is heavier, perhaps switching sides to control for our hand-dominance; drop a rock, or shout, down a well to judge how deep it is; or bounce a squash ball to estimate if it is warm enough for play. A key aspect of such behaviors is that they combine an intuitive understanding of how the physical world works, with actions that exaggerate, isolate, or bring into sharper relief a particular physical property of interest. Sometimes we perform these everyday experiments ourselves, while at other times we learn from observing others performing similar actions. Often, success in these endeavors is contingent on acting (or watching someone act) in appropriately "experimental" ways.

In this paper, we investigate how people learn about latent physical properties when interacting with virtual "microworlds". The microworlds are simulated environments on a computer screen. In these worlds, objects' movements are determined by a physics engine that approximates real-world physical laws (Ullman, Spelke, Battaglia, & Tenenbaum, 2017). In our experiments, we allow participants to freely grab and move the objects in the microworld. Our classification and quantification of their action strategies gives new insight into how people decide to act in the physical work to reveal information. Although it seems intuitively obvious that people engage in systematic behaviors during learning, it remains unclear how they decide which strategies to invoke, how complex they are, and how informative they are compared to other things they could have done. We presume people are effective because this matches out intuition, but it is an important scientific question to formally quantify and describe these abilities.

While active inference and physical exploration can be studied in naturalistic settings (Hoch, Rachwani, & Adolph, in revision; Kretch & Adolph, 2017; Piaget, 1936; Stahl & Feigenson, 2015), it is difficult to accurately measure and control all aspects of natural environments, and to measure participants' actions at a fine grained level. The current studies are designed to leverage some of the unique advantages of observing learning behavior in a virtual environment while also exploring a setting that inherits some of the complexity and dynamics that make the real world a challenging learning domain. Our virtual environments allow us to (1) precisely record and reconstruct every aspect of a learners' interactions and everything else that occurred during a trial (Rieber, 1996); and (2) develop formal models to quantify and objectively evaluate the information content of people's actions. Using this approach we are able to analyze in detail the types of actions people decide to use and how much information they generate for a given goal, helping to better understand the nature of our intuitive physical interactions.

The paper is structured as follows. We first lay out a normative framework for inference about latent physical properties of dynamically interacting objects, and show how we can use this framework to assess the informativeness of actions. We then describe two experiments that compare the inferential accuracy of active learners (who exert control over the objects in the microworld) with passive learners (who simply watch a movie of the microworld without interacting) and yoked learners (who watch videos of a previous active learner's sessions). In addition, we categorize active participants' experimental strategies with the help of our model-based information measures.

To foreshadow, across both experiments we find that active learners use sophisticated control to create situations that are highly informative about target properties (the properties they are are incentivised to identify) while minimizing confounding information about other non-target properties of the worlds. We conclude by discussing the scope of our findings more broadly. Specifically, we discuss how physical active learning strategies might be discovered, reinforced across instances, and generalized across contexts; and discuss the important connections between spatiotemporally extended active learning and adaptive control (Broadbent, FitzGerald, & Broadbent, 1986; Guez, 2015).

Active learning in discrete versus continuous and dynamic environments

In studies of "active learning", people shape their learning experience through their own actions (Gureckis & Markant, 2012). To date, active learning has primarily been studied in situations in which the learner's goal is to differentiate between a relatively small number of discrete hypotheses, such as the Wason card selection task (Oaksford & Chater, 1994; Sperber, Cara, & Girotto, 1995; Wason, 1968), category rule learning (Gureckis & Markant, 2009) and games like "Guess Who" (Nelson, Divjak, Gudmundsdottir, Martignon, & Meder, 2014), "Mastermind" (Berghman, Goossens, & Leus, 2009; Best, 1990; Goodrich, 2009; Hofer & Nelson, 2016) or "Battleships"(Gureckis & Markant, 2009;

Markant & Gureckis, 2014b, 2012). In these scenarios, participants pick from a fixed set of possible actions or questions in service of a learning goal, usually with each action-outcome pair contributing independently to a set of evidence they can use to make judgments. A subset of this research, on “active causal learning”, studies how people infer the underlying causal structure of simple dynamic systems (Bramley, Dayan, Griffiths, & Lagnado, 2017; Bramley, Lagnado, & Speekenbrink, 2015; Coenen, Bramley, Ruggeri, & Gureckis, 2017; Coenen, Rehder, & Gureckis, 2015; Lagnado & Sloman, 2002, 2004, 2006; Steyvers, Tenenbaum, Wagenmakers, & Blum, 2003). In a typical setting, participants perform interventions that manipulate variables within a causal system, and subsequently observe the consequences of their actions on the other variables (Bramley, Dayan, et al., 2017). For example, Coenen et al. (2015) had learners test computer chips to identify which of several possible wiring diagrams correctly described them. They could test them by activating one of the three components and observing whether either or both of the other two components activated as a result.

Evidence for the utility of active learning in these studies is mixed. Many studies find active learners produce more evidence than would be available if they did nothing, or behaved randomly. But other studies have demonstrated cases where people take stereotyped, or heuristic actions (Bramley et al., 2015; Coenen et al., 2015), that can be systematically uninformative (Wason, 1968) or fail to reveal particular kinds of rules (Markant & Gureckis, 2014a) or structures (Bramley, Dayan, et al., 2017). Active learners who consider the wrong hypotheses might produce less relevant evidence than would occur naturally (MacKay, 1992). For example, Markant and Gureckis (2014a) had people learn about both one and two-dimensional category rules. When learning actively about a two dimensional category rule, many participants wrongly expected the rule to relate to a single dimension, and so varied their tests only on that dimension. This resulted in their gathering less diagnostic information and performing worse at test than those who were exposed to a random selection of tests.

Yoking participants to the actions of another provides an additional window on active learning, separating the information that is in principle available, from the process of coming up with tests and updating beliefs. Active learners often outperform their yoked counterparts. One explanation for this result is the better match between the hypotheses active learners have in mind at any given time and the evidence they have to work with (Markant & Gureckis, 2014a). However, choosing how to act can be cognitively taxing, potentially overwhelming the advantages of active selection (Huttenlocher, 1962).

The idealized settings that past active learning research have focused on remove many of the inherent complexities of the real world. The physical world is dynamic, continuously changing, and the complex interactions between physical objects can mean their latent properties are rarely revealed unambiguously. Causal interventions in the real world extend across time (Bramley, Mayrhofer, Gerstenberg, & Lagnado, 2017) and space (cf. Wu, Schulz, Speekenbrink, Nelson, & Meder, 2017) in complex ways, and are grounded in and constrained by physical laws. For instance, the attraction between magnetic objects depends inversely on their proximity, a dropped rock's ability reveal the depth of a well depends on its falling cleanly and making an audible splash, and a medicine's effects on health are typically gradual, time lagged, and masked by other ongoing health factors.

As such, real intervention choices are much more unconstrained than those studied thus far in psychology. That is, the learner must not only choose *where* (i.e. on which variable) to intervene, but also, *how* and *when* to intervene — i.e. planning how to arrange and move objects to explore the nature of a magnetic field, or how to launch a rock so it will fall down a well without hitting the walls, or how much medicine to take and when. It becomes an interesting topic for psychological science to understand how people solve this action selection problem. One hypothesis we explore here is that the constraints come from the information momentarily generated by the interactions with the system. Another major challenge is having the right expectations about what should happen under different assumptions about the unobservable properties of the world. How would the learner expect non-magnetic objects to behave, or those with different magnetic properties? How long should the duration be before the splash for a particular depth of well? How unwell did you expect to feel, had you not taken a medicine? Given the mixed empirical results about human active learning abilities in the simple experimental contexts discussed above, it is an open and unresolved question to what extent people can act informatively in radically more complex and naturalistic learning scenarios. Furthermore, given the taxing nature of naturalistic control and planning (Osman, 2011), another open question is whether active learners will over or under perform yoked observers who can focus on the evidence without having to plan or carry out actions themselves.

In the current paper, we explore active learning in a continuously dynamic environment which reflects the underlying laws of physics that govern the motion and interactions of objects in the world. From a causal learning perspective, we can think of physics as a set of rich constraints on the form of a causal model — where relationships between entities must have functional forms that respect the equations of motion, conservation of energy and so on. Specifically, we will explore a two dimensional setting, in which an environment contains a number of colored circular pucks on a surface surrounded by walls, similar to a billiard or air hockey table (cf. Figure 1). While restricting to two dimensions makes these scenarios simpler than a three dimensional world simulation, such "billiard worlds" (Fragkiadaki, Agrawal, Levine, & Malik, 2015) have proven a useful setting for exploring intuitive judgments about physics in purely observational settings (Fragkiadaki et al., 2015; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012; Smith, de Peres, Vul, & Tenenbaum, 2017; Smith & Vul, 2014; Ullman, Stuhlmüller, Goodman, & Tenenbaum, 2014). In general, a two dimensional idealization of physics can readily be extended to the third because the equations of motion are solved separately for each dimension (Bottema & Roth, 1979).

While our setting removes some of the complexities of observing and interacting in the real world, our virtual microworlds are substantially richer than the domains in which human active learning has been studied quantitatively in the past. Thus, the present studies take a step toward more ecologically realistic learning settings. In the following section, we review past work on intuitive understanding of physics and some recent research that helps motivate the current studies.

Intuitive physics

Early research into intuitive physics highlighted ways in which people's understanding of some aspects of physics, such as ballistic and pendulum motion, is systematically biased (e.g. McCloskey, 1983). For example, when asked to draw the path of an object falling from a moving plane, many people will erroneously draw a line traveling straight downward rather than a parabola that correctly combines the initial forward motion with gravitational acceleration. This work suggests that physical understanding is often heuristic and context specific. However, more recent research has argued that some of these biases may be accounted for as resulting from optimal statistical inference assuming (1) our physical understanding is approximately Newtonian, and (2) we are often fundamentally uncertain about some important aspects of the physical scene (e.g., the masses of the objects involved in a collision, Sanborn, Mansinghka, & Griffiths, 2013).[1] For example, Battaglia, Hamrick, and Tenenbaum (2013) have argued that people's understanding of physics is best understood in analogy to a physics engine — a program that simulates physics to produce realistic scenes in movies and computer games (see also Ullman et al., 2017). Assuming that people have an approximate physics simulator in their mind helps explain how they can make predictions about what will happen in the future, reason about what happened in the past (Smith & Vul, 2014), or simulate what would have happened if aspects of the situation had been different (Chater & Oaksford, 2013;

[1]This perspective does a better job of accounting for biases relating to ballistic motion than angular momentum such as the tendency to predict a curved path for a ball exiting a curved tube (Kaiser, McCloskey, & Proffitt, 1986).

Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2015; Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum, 2017). For our purposes, the statistical inference approach provides a computational framework for understanding how inference about physical properties can proceed from observation of physical dynamics.

Learning about Latent Physical Properties

The relevant evidence for learning about physics is not the state of the world at a particular point in time, but rather how this state changes and evolves across time. Physical objects are pulled by forces, collide, and slide past one another in ways that depend on the laws of motion, the objects' latent properties such as mass and friction, as well as their susceptibility to forces. Given an understanding of the physical laws, it is possible to infer the latent properties of the objects involved in a scene from observation of sufficient dynamics. However, in order for learning to succeed, the right kind of dynamics have to be experienced. For instance, intuitively, we learn much more about the contents of a box that falls down a flight of stairs, compared to a box that sits at the back of our moving car.

Ullman et al. (2014); Ullman, Stuhlmüller, Goodman, and Tenenbaum (to appear) explored human inference about latent physical parameters from observing physical dynamics in 2D "microworlds" similar to the one shown in Figure 1a. In their setup, the worlds were bounded by solid walls and contained a number of colored pucks with differing masses, surfaces with differing levels of friction, as well as local (magnet-like) forces between pucks and a global (gravity-like) force pulling all the pucks in a particular direction. The properties of the worlds (the number and nature of the pucks, friction patches and forces) were generated from an underlying probabilistic program capable of generating a very large number of possible worlds. Participants watched and then replayed a five second clip from each of the generated worlds. In each clip, the pucks bounced around, attracting and repelling each other, being slowed down by the friction, and being pulled by the global force. Participants then answered a series of questions about each world's properties.

Ullman et al.'s (to appear) participants were able to detect different levels of mass and friction on average, but individual judgments were noisy. They identified the correct global force around 70% of the time and were much better at detecting local attraction (82%) than repulsion (53%). Ullman et al. found that this divergence was matched by an asymmetry in the evidence: pucks that repelled one another would rarely spend long enough close together to exhibit strong repulsion, while attracting pucks would rapidly approach one another and stick together offering better evidence of the latent force. Our experiments extend Ullman et al.'s (to appear) findings to a task in which participants

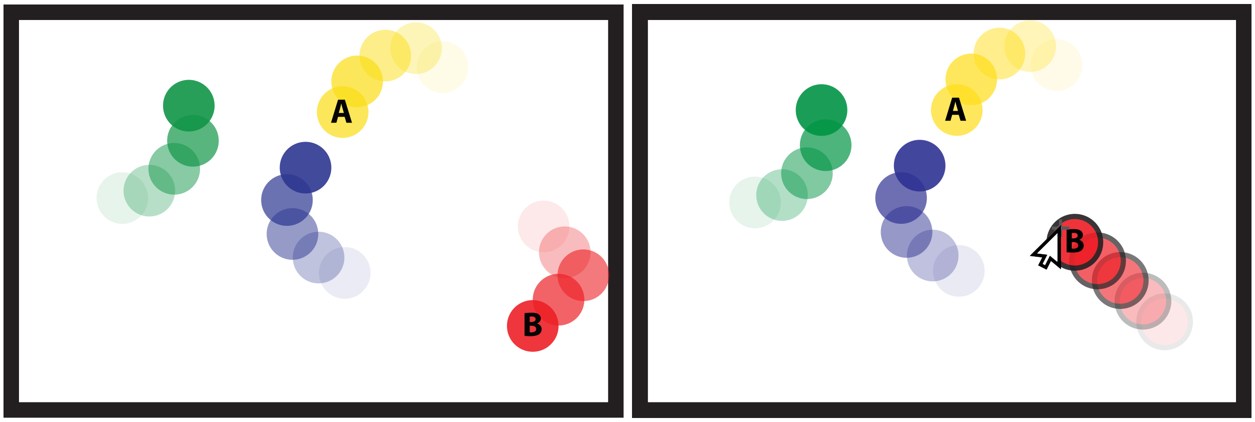


Figure 1. Schematic display of "microworlds" used in experiments. a) Four pucks are moving around colliding and affecting one another with local (magnet-like) forces. b) An active learner drags the puck labeled "B" (by left-hold-clicking on it and moving the computer mouse).

interact with the worlds to learn about the latent properties of the objects.

The Present Studies: Active Learning of Latent Physical Properties

Our task is adapted from Ullman et al. (2014, to appear). However, rather than preselecting scenes to show participants passively, we generated the scenarios dynamically during our experiments. This allowed us to include *active* conditions in which participants exerted control over the scenes that alter how they play out. We allowed active participants to grab objects by clicking on them and then drag them around using the mouse (see Figure 1b). We additionally included a *passive* condition in which participants merely observed the world (similar to Ullman et al.), and a *yoked* condition in which participants passively observed the actions of an active participant (described in more detail below). In our experiments, we chose to focus on how people learn about two target properties: local pair-wise forces, and object masses. We chose these properties because they were challenging for passive learners in Ullman et al's (2014; to appear) study, and because we hypothesized that inference about them might benefit from curated control.

Our primary goal is to establish whether people's actions are generally effective at reducing uncertainty about the specific parameters of the scene they are asked about. We are also interested in categorizing and making sense of the kinds of actions participants perform. In order to explore active learning quantitatively in this setting, we must first develop a method for inferring the latent properties of physical worlds based on observing and intervening with the objects in the world.

An Ideal-Observer approach to quantify inference from physical interactions

Like Ullman et al., we use an Ideal Observer (IO) approach to model inference in our task. The learner's goal is to infer the true latent properties of the objects. Essentially, our model works by predicting how the scene would unfold given different assumptions about these properties and comparing these predictions against observations. To the extent that the scene would play out differently depending on the value of some property, then observing what actually happens provides evidence about that property. For example, imagine determining what unknown substance is in an opaque box container (cf. Siegel, Magid, Tenenbaum, & Schulz, 2014). Simply lifting the box slowly might not reveal much about the contents (and thus be uninformative). Instead, shaking the box or tilting it may result in kickback from the liquid or sloshing sounds which help better narrow the possible materials. On the other hand, lifting the box gives good information about the weight or mass. These examples highlight how our everyday interventions with the physical world can vary in informativeness for our particular goals, such as determining either weight or the material content within the box. In the current context, we would expect an object to react differently to another object passing close by depending on whether they attract or repel one another and expect lighter objects to be moved more easily by attraction than heavier ones (cf. Figure 2).

The IO analysis allows us to assess how much evidence about the properties of interest is produced by the dynamics that a learner observes or brings about themselves. In particular, we can assess what participants' interventions reveal about different properties, and contrast this against the evidence that occurs "naturally" from passively observing what happens. While we focus on the tasks of identifying the masses of different objects, as well as their local force relationships (i.e., whether pairs of objects attract or repel each other), in principle any of the latent physical parameters can be inferred in this way, potentially even the laws of physics themselves (cf. Goodman, Ullman, & Tenenbaum, 2011; Ullman, Goodman, & Tenenbaum, 2012).

Inference. Given a video sequence of objects moving in the microworld (i.e., the data $\mathbf{d}$), the IO model uses simulation to infer the likelihood of possible world-settings $w \in \mathcal{W}$ that can be used to update a prior belief about the world settings $P(W)$ to a posterior $P(W|\mathbf{d})$ (where W is a random variable assigning a probability to all $w \in \mathcal{W}$). In our tasks $\mathcal{W}$, contains a world for all combinations of values of properties that are varied between different trials. These are the masses of the (labeled) target objects, and local forces relating the four objects (Figure 1). Our model assumes that people know how to apply the correct theory of physical dynamics to simulate how the world would unfold but lack knowledge of some of the parameters needed to fully specify the simulation. We will also assume knowledge of a number of parameters that are constant across all the worlds we consider (friction, elasticity, air resistance).[2] These modeling choices render the inference problem tractable and are reasonable given that (1) our simulator produces realistic dynamics, (2) participants learn about the invariant properties during the practice trials, and (3) we include a noise parameter that captures any additional sources of imprecision in both observations and simulations.

The basic intuition is that world dynamics (i.e., sequences of video frames) that are similar to the simulated predictions under a particular world setting w suggest that the true world is w. Divergence between a candidate world setting's predictions and reality can be measured throughout periods of observation and interaction. These divergences can be converted into likelihood scores assigning a probability of observing the actual object trajectories $\mathbf{d}$ given the potential world-setting $w \in \mathcal{W}$, the learner's interventions $\mathbf{c}$, and some Gaussian perceptual noise. Ullman et al. (to appear) based their inference model on divergence in absolute x, y positions of the objects. However, Vul, Frank, Tenenbaum, and Alvarez (2009) found evidence that human motion inferences reflect more sensitivity to motion information than location information, that is divergence in magnitudes r and directions θ of the objects' motion vectors. Thus, in this paper, we follow Vul et al. (2009) calculating the likelihood of each object's trajectory as

$$p(\mathbf{d}|w, \beta; \mathbf{c}) = \prod_{t=1}^{T} e^{-\frac{\beta}{2}(\mathbf{s}^t - \mathbf{d}^t)^\top \Sigma^{-1}(\mathbf{s}^t - \mathbf{d}^t)}, \tag{1}$$

where

$$\Sigma = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_\theta^2 \end{bmatrix}. \tag{2}$$

$\mathbf{d}^t$ is the object's actual $[r, \theta]$ motion at time t, while $\mathbf{s}^t$ its simulated motion if w is true.[3,4] To ensure that both the divergence in magnitude and direction of velocity affect the likelihoods to a comparable degree, we set σ_r and σ_θ to the empirical standard deviations of the divergences between a random sample of the simulations and the observations on these dimensions. Finally, β is a convenience scaling parameter capturing, roughly, a learner's presumed imprecision estimating the objects' actual and simulated magnitudes and directions. The key property of Equation 1 is that smaller divergences between a

[2] In Box2d, friction governs how objects slide past one another when in contact and air resistance governs how they lose energy while moving without touching anything.

[3] For divergences in angle θ, we always assume the shortest route around the unit circle, i.e. divergences are always $> -\pi$ and $< \pi$.

[4] In the Appendix, we examine a range of alternative distance measures based on combinations of both location and motion information.

simulation and observation lead to higher likelihoods. Figure 2a provides a visualization of this procedure.

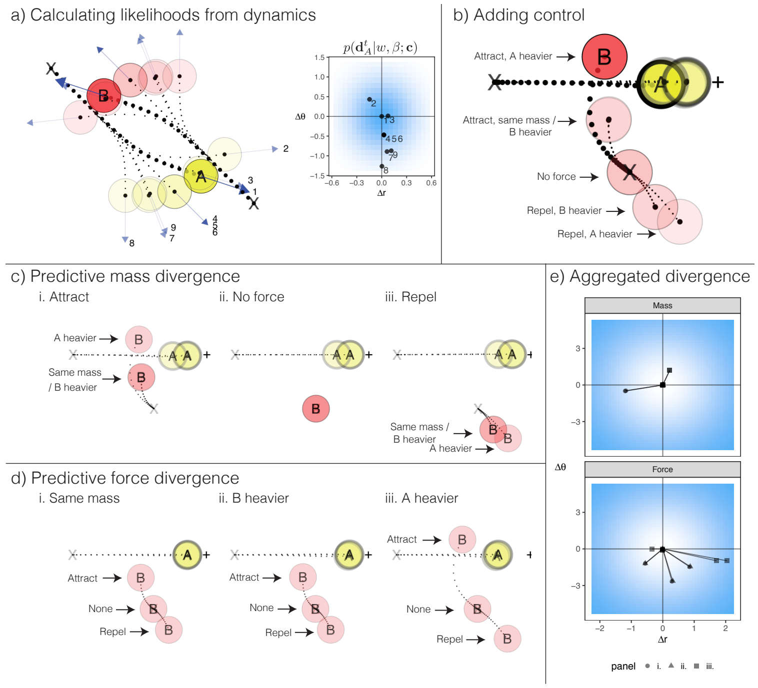


Figure 2. a) Calculating likelihoods from observed dynamics. A and B attract one another and have the same mass. *Observed paths* originate at “X” and follow thick dotted lines. *Simulated paths* for different masses and local forces follow thin dotted lines. The simulation with the true world settings lies on top of the observed path (attract, same mass). Likelihoods are based on the divergence between simulated and observed trajectories (blue arrows). The Likelihoods for the simulated positions of A under $w \in 1:9$ (numerical labels) are shown in the plot on the right. b) Example involving active control. A is dragged to the right by the active learner moving close by B (“+” symbol indicates final position of cursor). Here, A and B attract each other and A is heavier. Simulations in which A and B repel each other predict qualitatively different movement for ball B. c) PD_{mass} contrasts simulations on mass property, averaging over other properties. Here we aggregate over force: left, middle and right in the example from b). d) Shows the same contrast for PD_{force}. e) Visualizes the divergences from c) and d) in terms of in terms of r and θ. PD_{mass} and PD_{force} are the grand means of the black divergence lines in the top and bottom panels respectively.

There are, however, two complicating details. First, the physical dynamics involving multiple bodies, local forces and collisions are notoriously chaotic (e.g. Bruns, 1887). This means that simulations are likely to diverge dramatically from observations over time unless they have exactly the correct parameters and starting conditions (i.e., only the exactly right simulation will resemble the world over a long time interval). A consequence of this is that long simulations without correction provide poor signal as to whether one setting is closer to the truth than another. We address this issue by focusing on instantaneous divergence. Concretely, we continually return simulated objects to their observed trajectories and allow them to start diverging anew. Repeating this throughout a learning period and integrating across time provides a continuous measure of the likelihood of world dynamics under any specific world settings.[5]

The second complicating detail concerns the role that active interventions have on the physics within the microworld. Our approach incorporates active learning into this scheme in analogy to the causal Bayesian network treatment of interventions as "graph surgery" (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). The idea is that interventions are actions that affect the dynamics going forward, but that are not themselves *caused by* the preceding dynamics. For example, in Figure 1b the participants' mouse movement affects "B"s trajectory directly, which has resulting effects on the other objects. Thus, interventions can be thought of as "acts of God", or "little miracles" (Lewis, 1973) that have no causal explanation within the world.[6] This means we can condition on the learners' actions when interpreting the evidence produced by interactions with the world (see Figure 2b).

The likelihood of a world given dynamics $\mathbf{d}_{1:T}$ and interventions $\mathbf{c}_{1:T}$ is assumed to be the product of the likelihoods of all the divergences measured at every "instant" $t \in T$ throughout a learning period. By computing these likelihood scores for all possible world-settings and combining with a prior $P(W)$, we can compute a posterior over worlds $P(W|\mathbf{d}, \beta; \mathbf{c})$ and associated posterior uncertainty $\mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c})$ using standard Shannon entropy (Shannon, 1951):

$$\mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c}) = -\sum_{w \in \mathcal{W}} P(w|\mathbf{d}, \beta; \mathbf{c}) \log_2 P(w|\mathbf{d}, \beta; \mathbf{c}). \tag{3}$$

$\mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c})$ provides a scalar measure of the degree of remaining uncertainty about the true world's latent properties at the end of each trial. This measure depends on what interactions occur during a clip. For example, the trajectories of two objects are only (strongly) affected by local forces if the distance between the objects is small at some

[5]Note that in our simulations we hold fixed the size of the time windows over which the instantaneous divergences are measured at 10 frames and do not fit the noise parameter, always assuming $\beta = \frac{1}{50}$. Thus, the resulting measures should be thought of as a guide to the relative rather than absolute evidence available in a trial. $\beta = \infty$ corresponds to perfect knowledge of the objects' locations and velocities, which, combined with perfect knowledge of the physics engine, rules out all but the true world-settings within a few frames. $\beta = 0$ would assign equal likelihoods for all worlds regardless of the evidence.

[6]Of course, the learner's actions are likely to be influenced by the dynamics observed prior to acting together with the learner's goals and beliefs. Extended interventions in dynamic systems will also likely involve motor feedback and corrections. We do not model these complexities but will return to them in the General Discussion.

point. And objects need to collide with one another (or be dragged by the mouse) for their trajectories to be dependent on their masses.

We express the cumulative value of a period of observation and intervention as the ideal observer's reduction in uncertainty

$$\Delta \mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c}) = \mathrm{H}(W) - \mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c}). \tag{4}$$

Within this framework, we can calculate posteriors over particular parameters of interest (such as the local force between a pair of objects) by marginalizing over the remaining parameters. We will use this procedure to assess what evidence participants generated through their actions, as well how much evidence was present in the passive dynamics of a clip.

Assessing the informativeness of actions: Predictive divergence. We use the IO model not only to evaluate the information available in any display over an extended learning period, but also to determine the quality of specific actions in terms of how much information they convey about the relevant latent properties. This is related to the box example above, where different actions reveal more or less information about the contents of a box.

We created a novel measure called Predictive Divergence (PD). PD captures the extent to which what happens at a given moment in the scene *depends* on a particular property. We calculate PD for a given property by simulating the world forward, then taking the divergence between simulations that vary on that property and what actually happened (e.g., whether A attracts, has no effect on, or repels B) while also averaging over all possible settings of the other properties. The result is a measure of how strongly a property of interest is revealed at any point throughout a trial (see the Appendix for details).

As a concrete example, consider Figure 2b–d. All these subplots show the same action: object A is dragged past object B. Figure 2c visualizes the extent to which the simulated trajectories depend on the objects' mass ($\mathrm{PD}_{\mathrm{mass}}$). This shows that, regardless of the force relation, how quickly A is dragged to the right depends a little on its mass. If A is heavier, it is dragged more slowly compared to when A is lighter. Additionally, if the objects attract (left panel) or repel each other (right panel), this affects how far B will travel. Figure 2d visualizes how the objects would move for different local forces. How B travels differs substantially depending on whether there is an attractive or repulsive force between A and B and this is true whatever the masses are.[7] Figure 2e visualizes the

[7]One might wonder why the left and center panels of Figure 2d look so similar. Since B is heavier in the

resultant PDs in terms of differences in r, θ. In both cases PD is the grand mean of the black divergence lines, showing that this action is more informative about force than mass.[8]

To a first approximation, generating high PD with respect to a target property is a good objective for planning interventions. Interventions with high PD correspond to situations where one can expect very different things to happen depending on the truth about the target property. In parallel, when an intervention also has low PD for other properties, these expectations are less confounded by the learners' current uncertainty about these other properties. For example, in Figure 2d, the fact that the force effects work out similarly regardless of the relative mass (panels i., ii., and iii. are similar) means that a learner can make a strong judgment about the force based on this action even if they are still very uncertain about the relative masses.

Unlike uncertainty reduction, PD does not depend on the learner's current prior. Because we take an unweighted average over possible settings of the other parameters of the world, the PD is naturally agnostic about what these parameters might be. This is useful for the current context where we are interested in how informative different actions are in general, not relative to specific parameters of the aspects of the world. PD also differs from our calculation of likelihood by being prospective. Likelihood calculation involved a retrospective comparison of observed against simulated trajectories, whereas PD involves a prospective comparison between different simulated trajectories (cf. Figures 2). This makes it a useful measure for planning efficient actions similar to models of information gain in the optimal experiment design literature (Coenen, Nelson, & Gureckis, submitted).

Overview of Experiments

In our experiments, we focus on how people learn about two target properties: local pair-wise forces, and object masses. We created scenarios where evidence is frequently confounded by including two "distractor" pucks along with two "target" pucks (whose properties participants are asked to infer) and drawing local forces randomly for all pairs of target and distractor pucks. This means that it is important to isolate the target pucks from the distractor pucks to get clear information about the target properties.

Both experiments feature between-subject manipulations that contrast active, passive, and yoked learning (in which a passive learner observes the actions of an active

middle panel, we might expect it to move less. However, the mouse's influence dominates A's path because it increases with distance from the cursor (as if the object is attached to the cursor by an elastic band) while B's influence decays with distance (like a magnet's). Since A's path is not deflected substantially by B, B will travel toward or away from A in this situation almost irrespective of its mass, similar to how objects accelerated toward Earth accelerate at $\approx -9.8\,{}^{\mathrm{m}}/{}_{\mathrm{s}^2}$ no matter how how heavy they are.

[8]As with the uncertainty calculations, we examine a range of distances for driving the PD calculations in the Appendix.

learner). Comparing active to passive learning lets us assess whether learners' actions increase or decrease the information about the target properties with their actions or decrease it, and whether this results in changes in inferential accuracy. Comparing active learners with yoked learners is common in the active learning literature and offers additional insight into the learning mechanisms at play (e.g. Markant & Gureckis, 2014a). In Experiment 1, participants had two inference goals on each trial: to determine 1) the local force between the target objects, and 2) their relative masses. Comparing passive and active learners here allows us to identify differences in the evidence produced in either case. Meanwhile, comparing active against yoked learners equates the evidence that is in principle available, allowing us to assess differences in judgment that might stem from whether the learner has control over what happens. In Experiment 2, active learners were given a single learning goal per trial and we manipulated whether yoked participants had the same or a different goal. This allows us to assess the extent to which active learners produce evidence that is *specific* to their learning goal, and how having a matched or mismatched learning goal affects the performance of yoked learners.

We identify a range of active learning strategies performed by active participants in Experiment 1, and use our framework to assess how each strategy helps achieve the different learning goals. In Experiment 2, we then assess these claims empirically, contrasting how frequently these strategies are performed dependent on the active learners' goal.

Predictions. The key hypothesis is that active learners will tailor the information they produce during a trial to be revealing about the latent properties they are asked to infer. This should give an advantage to both active and yoked participants (simply because they have better data than the passive participants who lack control).

Participants and the inference models considered in Ullman et al. (2014, to appear) had difficulty with identifying repulsion in particular. As mentioned in the introduction, this is because objects that repel one another rarely pass close enough together to exhibit this force in naturally occurring dynamics. We expect to replicate this effect in our passive condition both in terms of information produced and accuracy of inferences. However, we predict that active learners will show a different pattern, since they can counteract the information asymmetry by bringing the target pucks closer together and thus produce strong evidence about repulsion.

It is frequently the case that yoked learners underperform relative to active learners, possibly reflecting a learning advantage for volitional control (e.g. Markant, Dubrow, Davachi, & Gureckis, 2014; Markant & Gureckis, 2014a). Given the complex nature of the control and inference in the current task, we hypothesized that active participants would outperform yoked participants. Additionally, in Experiments 2, we expected yoked participants whose goal was different than their active counterparts' to perform less accurately because the active participants were expected to produces little evidence relevant to the yoked participants' goal.

Experiment 1

Methods

Participants. Sixty-four participants were recruited from Amazon Mechanical Turk (39 male, $M \pm SD$ age 33.6 ± 10.2) using psiTurk (Gureckis et al., 2016). The number of participants was chosen by a heuristic goal of running a minimum of twenty participants per condition and no statistics were performed until the entire data set was collected. Participants were paid at a rate of \$6 per hour, plus performance-related bonuses (\$$0.61 \pm 0.17$).

Conditions. Participants were pseudo-randomly assigned to one of three treatment conditions:

1. **Passive ($N = 24$)** Participants observed the microworlds unfold without being able to interact. If, in rare cases, everything came to a standstill, objects' locations and initial velocities were reset.
2. **Active ($N = 20$)** Participants could grab pucks and drag them around with the mouse. Grabbed pucks retained their properties (i.e. mass, local forces, location, and momentum) but became strongly attracted to the position of the mouse, as if attached by an elastic band. While it is impossible to completely replicate the haptic feedback inherent to real-world physical actions, this choice allows objects to react to control in ways that depend on their properties. That is, heavier objects accelerate slightly less than lighter objects when "pulled" with an equivalent mouse motion, and local forces compete with attraction to the mouse in determining a controlled object's trajectory.[9]
3. **Yoked ($N = 20$)** In this condition, each participant was yoked one-to-one with an active participant, and watched replays of the active participant's interactions on each trial. As with the other conditions, they watched each replay once through, without stopping.

[9]We opted for strong attraction rather than simply copying the position of the mouse because this allowed the controlled object to interact reciprocally with the other objects in collisions. Otherwise, controlled object would have one-sided interactions with the objects with which it collided, behaving as if it was infinitely heavier than them. This resulted in unrealistic and chaotic interactions when the controlled object was moved fast or trapped another object against a wall.

Worlds. Each participant watched or interacted with nine microworlds, consisting of all combinations of target force in *attract*, *repel* and *none* and target masses in $[1,1]$kg, $[1,2]$kg and $[2,1]$kg (see Table 1).[10] Each world also had up to five additional local forces, one between the target pucks, and one for every other combination of target and non-target puck. These were drawn uniformly from the three possibilities for each participant on each trial and the two distractor objects always weighed 1kg. This results in an overall set of 2187 possible worlds $w \in \mathcal{W}$ (e.g. all 3^7 combinations of target and distractor local forces and the possible target masses) but a smaller judgment space containing the nine combinations of target mass and target force. The settings for all other properties of the objects (elasticity, friction etc.) were the same for all worlds, as detailed in Table A1 in the Appendix.

Table 1
Experiment Design

World	1	2	3	4	5	6	7	8	9
Target force	A	A	A	N	N	N	R	R	R
Target 1 mass	1	1	2	1	1	2	1	1	2
Target 2 mass	1	2	1	1	2	1	1	2	1

Note: A = attract, N = none, R = repel; masses are in kg.

Materials and Procedure. The experiment was programmed in Javascript using a port of the Box2D physics game engine. Open source demos of all three conditions and replays for all participants and trials are available at our online repository (`https://github.com/neilbramley/active_physics`). A complete specification of the settings of the Box2D simulator is available in the Appendix.

On each trial, the initial position of each puck was random but non-overlapping with initial velocities in the x and y direction drawn at random. If, in the passive condition, all pucks' came to a near standstill, the simulation froze and the window went black briefly before the positions and velocities of the pucks were redrawn.[11] Each world was simulated for 45 seconds at 60 frames per second, leading to 2700 frames of evidence per trial.

The microworlds were displayed in a 600 by 400 pixel frame, with 1 m in the world corresponding to 100 pixels on the screen. Each world was bounded by solid walls with high elasticity – and contained four pucks of random colors that were different on each trial.[12] The two target pucks were labeled with new letters on each trial (e.g. "A" and "B"

[10] Local forces scaled with the inverse squared distance between the objects in line with Newton's universal law of gravitation. Thus, the current local force L exerted on object o_1 by object o_2 (and the reverse) was given by $\pm 3\frac{m^1 m^2}{d^2}$.

[11] This happened 1.0 ± 0.83 times per 45 second trial on average.

[12] To minimize the possibility that participants might group the pucks based on their colors or transfer

on trial one, "C" and "D" on trial two, cf. Figure 1) while the distractor pucks were unlabeled. This was done to minimize transfer effects and confusion between the objects in the different trials which had been an issue in Bramley et al.'s (2016) pilot experiment. For yoked participants, the cursor of the active participant was shown with a "+" symbol whenever it was within the world, and any objects grabbed by the participant were indicated as in the active condition with a thick black border (see Figure 1b).

At the end of each trial, our two test questions appeared in counterbalanced order below the world. The mass question asked "Is A or B heavier?" and participants responded with "A", "same", or "B". The force question was "What is the relationship between the pucks marked A and B?" and participants responded "they attract each other","none", or "they repel each other". The puck labels in the questions changed between trials as mentioned above. To ensure that participants were motivated to accurately answer the questions, we paid a 5¢ bonus for each correct response. Participants also gave a confidence judgment for each question, indicating "How sure are you that you got this question right?" using a 100 point slider ranging from "not at all" to "very much".

Participants first completed instructions relevant to their condition, answered comprehension check questions, and then completed two practice trials followed by the nine test trials. Practice trials were always worlds one and five shown in Table 1.[13] Practice trials were indistinguishable from test trials from participants' perspective and were excluded from analysis. At the end of the experiment, participants received feedback about how many of the test questions they got right and the bonus they had earned. The experiment took 19.0 ± 7.3 minutes to complete.

Results

We first look at participants' inferences about the latent properties, before exploring the information produced by the action in the trials. For judgments, we contrast accuracy by condition, asking whether active learners outperform passive learners, and whether yoked learners inherit this advantage. We then contrast judgment patterns for the two types of test question (mass judgments and force judgments) assessing whether we replicate the same patterns for passive learners as Ullman et al. (2014; to appear) and crucially, whether active control affects these accuracy patterns.

Accuracy. Overall accuracy, including both question types, differed significantly by condition $F(2, 61) = 3.8, \eta^2 = .12, p = .03$ (see Figure 3). Active participants answered

their properties erroneously across trials, the four pucks' had hues equally spaced around the HSL color wheel with a new random starting point on each trial.

[13]Note that worlds one and five were also presented in the test phase. However, the randomly drawn distractor forces, puck colors, and labels differed between the practice and test instances.

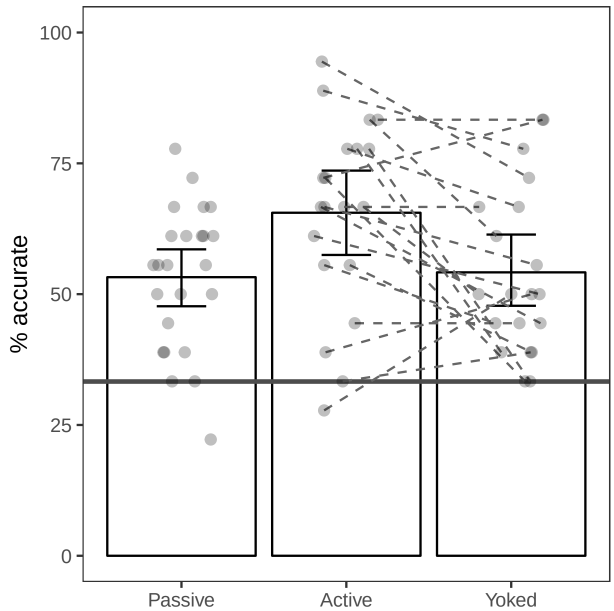


Figure 3. Experiment 1: Performance by condition measured by the accuracy with which participants' correctly answered the three-alternative forced-choice questions about the local forces and object masses. Bars denote condition means. Error bars denote bootstrapped 95% confidence intervals. Points denoting individual participants are jittered along the x-axis for visibility. Dashed lines connect active participants with matched yoked participants. The horizontal line indicates chance performance.

significantly more questions correctly than passive participants, $t(42) = 2.5, p = 0.014$, and their yoked counterparts, $t(19) = 2.9, p = 0.01$, with no difference between passive and yoked participants, $t(42) = 0.2, p = 0.83$. Active participants' performance was predictive of their yoked counterparts', $r = .49, t(18) = 2.4, p = .03$.

Confidence judgments differed by condition, $F(2, 61) = 5.3, \eta^2 = .15, p = .007$, with active participants (M $\pm$ SD: $78.5 \pm 13.5\%$) significantly more confident on average than passive $(66.7 \pm 13.9\%), t(42) = 2.8, p = .006$ and yoked $(65.4 \pm 16.1\%)$ participants, $t(38) = 2.9, p = .006$, but no difference between passive and yoked $t(42) = 0.28, p = 0.78$. Overall, participants were significantly more confident about their correct 73.4 ± 16.0 than their incorrect 63.8 ± 17.2 judgments $t(63) = 6.4, p < .001$.

Mass versus force. On the mass question, participants answered $46 \pm 29\%$, $58 \pm 24\%$, and $45 \pm 20\%$ of questions correctly in the *passive*, *active* and *yoked* conditions respectively. On the force question, participants answered $61 \pm 22\%$, $73 \pm 21\%$, and $63 \pm 21\%$ of questions correctly. Across conditions, participants were worse at inferring masses than forces $t(63) = 4.8, p < .0001$ and reported lower confidence in mass judgments $(66 \pm 25\%)$ compared to force judgments $(74 \pm 25\%)$ $t(63) = 4.2, p < .0001$. As predicted,

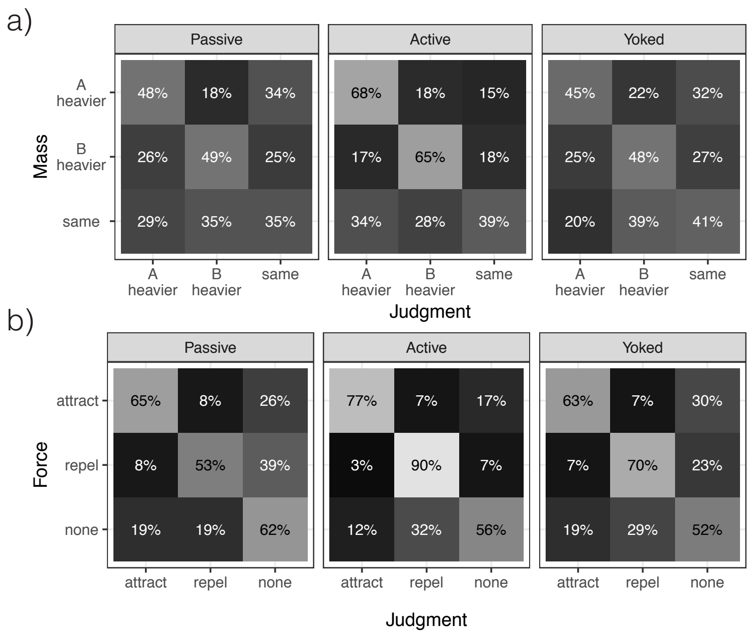


Figure 4. Experiment 1: Confusion matrices for mass question (a) and force question (b). For example, the mass question in the passive condition, left matrix in a), when A was heavier than B (top row), participants correctly identified that this was the case 48% of the time, they falsely judged that B was heavier than A in 18% of the cases, and thought that both A and B have the same mass in 34% of the cases.

active learners benefited particularly on trials in which the target objects repelled each other. Passive participants only identified repulsion correctly 53% of the time compared to 90% for active participants (see Figure 4). Force type interacted with treatment condition in predicting accuracy $F(6, 183) = 3.0, p < .0001$. Dummy contrasts with "no force" and "passive" as controls revealed that active participants were significantly better at identifying repel than passive participants $t(42) = 3.2, p < .0001$ and there was a marginal improvement for yoked participants as well $t(42) = 1.9, p < .058$. There was no significant relationship between accuracy on the local force question and the number of distractor forces.

Summary. As predicted, we found that active participants outperformed passive controls. In particular, they were better at identifying repulsion, consistent with the idea that they often pushed repulsive target objects close together. Active participants also outperformed their yoked counterparts. This suggests that volitional control was crucial to their successful use of the generated evidence. Together, these findings raise questions about *what* the active participants were doing, and *how* their actions helped them identify the worlds' properties. In the next section we begin to explore this both *quantitatively* by measuring the evidence generated throughout each trial, and *qualitatively* by categorizing the different testing strategies that active participants came up with.

Information

We now use our Ideal Observer (IO) model to better understand why active participants generally outperformed passive participants, and why active participants found repulsion easier in particular. We simulated each trial under all 2187 possible world settings and tracked how much they diverged from what actually happened, returning simulated objects to their actual trajectories every 10 frames (i.e., 6 times per second).[14]

We assumed learners began each trial with a uniform prior over worlds $P(W) \sim \mathrm{Unif}(|W|)$ and computed a posterior $P(W|\mathbf{d}, \beta; \mathbf{c})$ for every trial. We then calculated posterior uncertainty relative to the target mass and force questions by marginalization over this posterior. Because the accuracy of our model depends on what level of imprecision we assume (captured by β), we are not interested in comparing participants' and models' *absolute* accuracy. Instead, we assess whether the IO model's notion of *relative* evidence lines up with participants' judgments across questions and worlds.[15]

We find that posterior uncertainties over the full space of possible worlds did not differ on average for passive compared to active participants $t(29) = 0.47, p = 0.64$.[16] Nor did posterior uncertainty about mass $t(29) = 0.45, p = 0.66$ (Figure 5a). However, active participants generated significantly more information on average about the target force $t(29) = 3.8, p < 0.001$ (Figure 5b). Mass uncertainty appears somewhat bimodal for active participants, with over half of participants (7/13) achieving more certainty on average about mass than all 18 of the recorded Passive participants, but the other half doing a little worse than those passive participants on average. The probability of answering the mass question correctly was marginally inversely related to posterior mass entropy across all trials $t(394) = -2.0, p = 0.046$ while this did not hold for force information $p = 0.11$. Inspecting Figure 5, we also see that mass uncertainty was significantly lower than force uncertainty, on the order of 0.1 bits compared to 1 bit $t(60) = 8.1, p < .001$. This means the Ideal Observer was generally substantially more certain about the mass question, in contrast to participants who were less accurate at inferring mass.

In sum, while active participants generated about the same amount of evidence

[14]Like Ullman et al. (2014), we experimented with several "snap back" windows finding no systematic differences in the results.

[15]We found no systematic differences in results for different values of β. We chose $\frac{1}{50}$ simply to make the relative differences in posterior uncertainty easy to visualize in Figures 5 and 10.

[16]Because the objects' velocities were not recorded for 13 participants, information measures for these participants could not be computed. The following comparisons are based on the remaining 18 passive and 13 active participants (and the 13 corresponding yoked participants where appropriate).

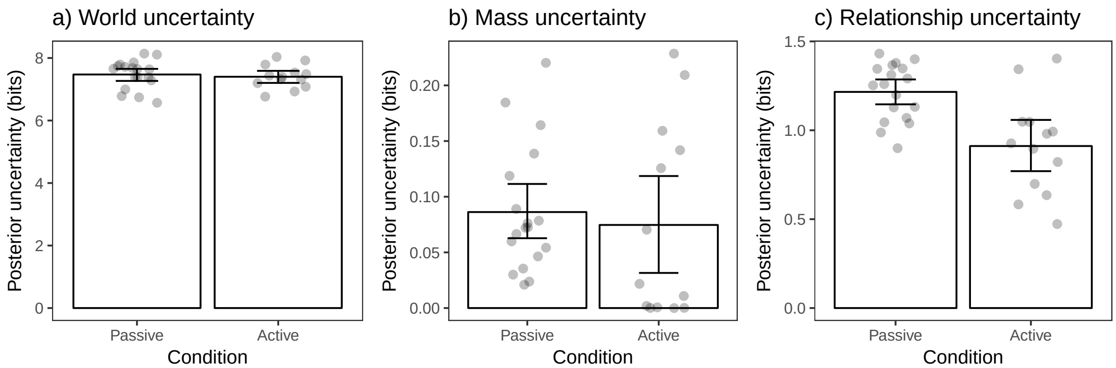


Figure 5. Experiment 1: Mean posterior uncertainty according to ideal observer model assuming noise parameter $\beta = 10$. a) Uncertainty about the true world e.g., $\mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c})$. b) Uncertainty about target masses e.g., $\mathrm{H}(\text{mass}|\mathbf{d}, \beta; \mathbf{c})$ (i.e., after marginalizing over other other properties). c) Uncertainty about the target forces $\mathrm{H}(\text{force}|\mathbf{d}, \beta; \mathbf{c})$. Bars denote condition means. Points denoting individual participants are jittered along the x-axis for visibility.

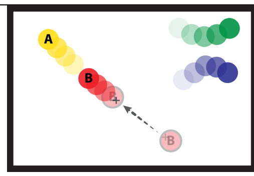
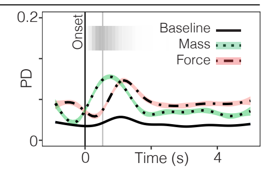
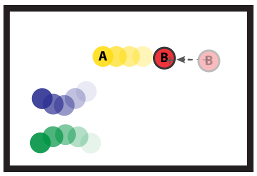
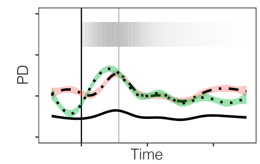
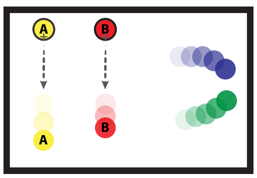
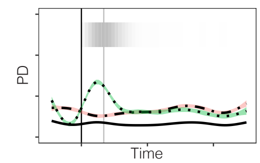
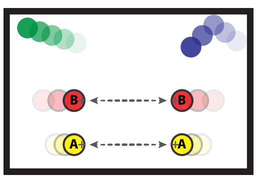
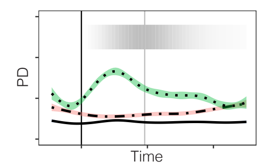
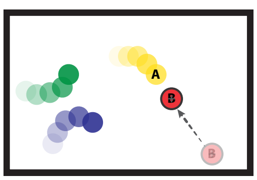
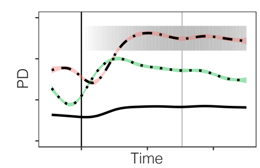
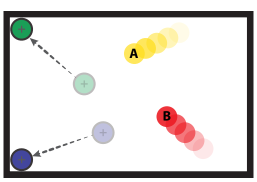
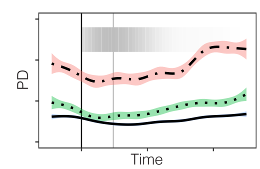
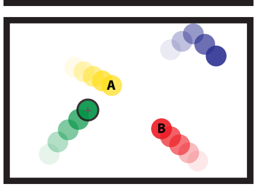
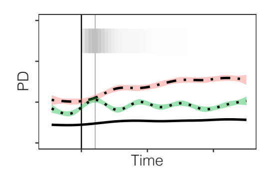
overall as passive participants (about all properties of the worlds) they generated more evidence about force than passive participants. Furthermore, while they did not generate systematically more evidence about masses, there was some evidence that the amount they did produce was consequential to judgment accuracy.

Experimental strategies

To get a better sense for what actions participants performed, we viewed the active participants' replays from Experiment 1 and identified a number of strategies. We describe these in Table 2. We also provide replays from the experiments that exemplify the different strategies in the online repository (`https://github.com/neilbramley/active_physics`).

To what extent did participants make use of the different strategies? Some of these proposed strategies have easy-to-measure hallmarks. For instance, in line with the *shaking* strategy (Figure 2d), participants who moved the controlled object around faster did better on the mass question $\beta = 25, F(1, 18) = 15, \eta^2 = .45, p < 0.001$, but this had no relationship with accuracy on the force question $p = .67$. Conversely, in line with *encroaching* (Table 2e), we see evidence that participants in the active condition identified the local forces by bringing the two target pucks close to each other. The lower the average distance between two target objects for an *active* participant, the better they did on the force question $\beta = -.3, F(1, 18) = 8.0, \eta^2 = .3, p = .001$ but this had no relationship with accuracy on the mass question $p = .87$. In Experiment 2 and the subsequent analyses we will explicitly link these strategies to our information measures.

Table 2
Strategies observed in Experiment 1.

Strategy	Schematic	Description	Profile
a) Launching		Grabbing one of the target pucks and “throwing it” against the other target puck.	
b) Knocking		Grabbing one of the target pucks and knocking it against the other (without letting it go)	
c) Throwing		Grabbing a target puck and throwing it, avoiding collision with any of the other pucks.	
d) Shaking		Grabbing a target puck and rapidly shaking it from side to side.	
e) Encroaching		Grabbing one target puck and moving it close to the other.	
f) Deconfounding		Grabbing a distractor puck and moving it away from the target pucks (e.g. into a corner).	
g) Controlling		Briefly grabbing a fast moving puck and releasing it to slow it down.	

Note: **Schematic**: *target pucks* are labeled A or B, and *distractor pucks* are unlabeled. **Profile**: Predictive divergence profiles for coded strategies in Experiment 2, smoothed using a GAM (Hastie & Tibshirani, 1990), with fills showing 99% confidence intervals. Black vertical lines mark onset of control. Shaded horizontal fills and gray vertical lines respectively indicate range and median time after onset at which the participant let go of the puck.

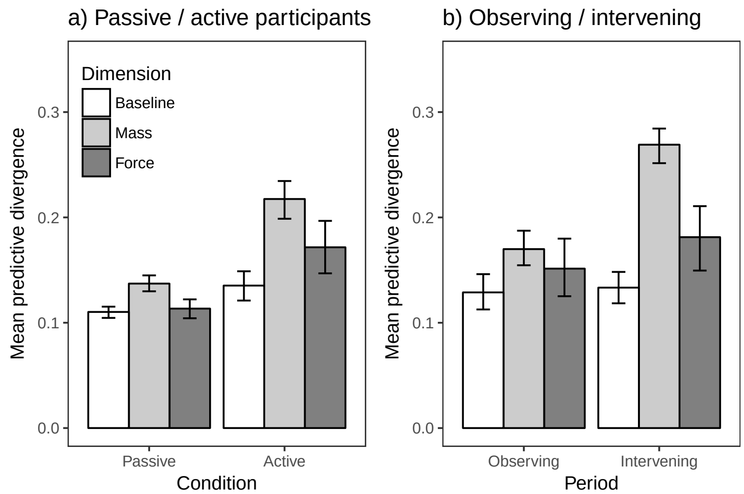


Figure 6. Experiment 1: Comparison of average predictive divergence for (a) passive and active participants, and (b) periods of observation and intervention for active participants. Bars denote means and error bars denote bootstrapped 95% confidence intervals. “Baseline” denotes the average PD measured for every dimension over which the worlds in $\mathcal{W}$ vary.

Predictive divergence. To get a better sense of what participants were doing, we applied the Predictive Divergence (PD) measure from the model (see *Assessing the informativeness of actions* in the Introduction). This allows us to assess the information available about the world's latent properties at different points during a trial. We considered three variants: PD_{mass} measures the current predictive divergence depending on the target objects' masses. PD_{force} does the same for the target force. We compare these measures against Baseline, which is the average at each time point of the predictive divergence for all the properties of the world that could vary across trials (the masses of the targets, their force relation but also the five distractor forces). In the passive condition, natural dynamics should not privilege the target over the distractor properties. This is shown in Figure 6a where PD_{mass} and PD_{force} are similar to baseline. In line with the fact that the ideal observer generally found natural dynamics to better reveal mass than force, PD_{mass} is a little higher than baseline $t(34) = 7.5, p < .001$ but there there is no difference between PD_{force} and baseline $t(38) = 0.89, p = 0.38$.

Active participants had higher average baseline $t(29) = 3.4, p = 0.002$, PD_{mass} $t(29) = 8.3, p < 0.001$ and PD_{force} $t(29) = 4.4, p < 0.001$ than passive participants (see Figure 6a). Additionally, looking within active learners' trials, we compared periods of active control to periods of passive observation. We found that periods of control exhibited no increase in baseline divergence $t(12) = 0.56, p = 0.58$ but a large increase in PD_{mass} $t(12) = 9.6, p < .001$ and a barely significant increase in PD_{force} $t(12) = 2.1, p = 0.0498$ (see

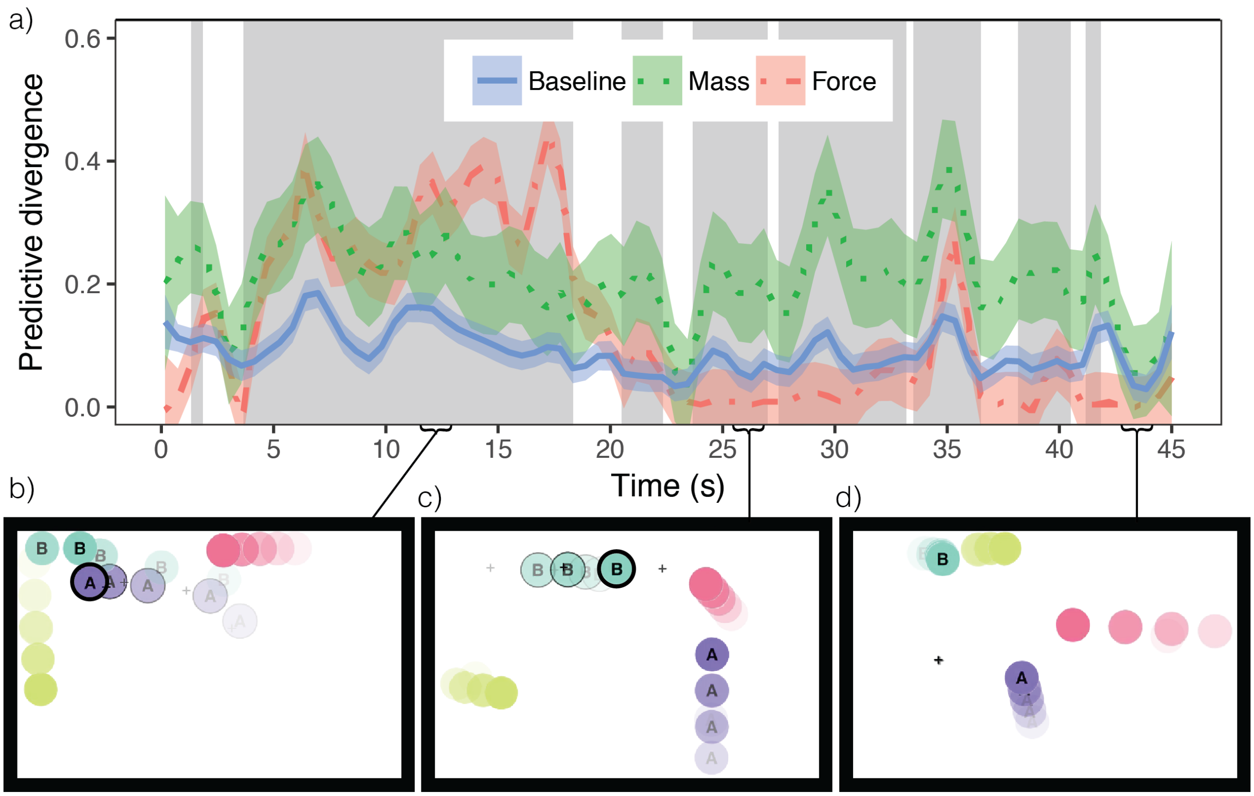


Figure 7. Example trial from Experiment 1. Replay: `https://neilrbramley.com/experiments/apl/e1/replays?p=1` a) Timeline for PD. Lines indicate PD for any dimension (solid), mass (dotted), and force (dot–dashed), smoothed using a LOESS kernel (Cleveland, Grosse, & Shyu, 1992) with fills indicating 90% confidence intervals. Gray rectangles indicate periods of control. b–d) Visualizations of actions during clip. A sequence of screen-shots are superimposed with later frames becoming more opaque. Thick black circles indicate controlled object and "+" indicates the mouse.

Figure 6b).

Figure 7a shows a timeline from an active participant's trial. At first, the participant takes control of puck A and brings it close to puck B (Figure 7b). He or she also knocks the targets together several times. This generates considerably more PD_{mass} and PD_{force} than baseline. In a later part, the participant takes puck B and shakes it back and forth (Figure 7c), and later takes puck A to do the same. These periods generate high PD_{mass} (because a heavier controlled object reacts more slowly to changes in direction) but not PD_{force} (because the target pucks are typically not close together). When the participant stops controlling, PD_{force} and PD_{mass} drop to close to baseline (Figure 7d).

Discussion

Experiment 1 revealed a benefit for active over passive learning in this dynamic physical setting. Participants' actions resulted in more information about force, and in some cases, more information about mass, too. In particular, active participants were able to gather more evidence about repulsion by bringing target objects close together and moving distractor objects out of the way. A number of learners also gathered information about masses by shaking the target objects back and forth, and generally staged many more interactions between targets and fewer between distractors compared to passive participants.

Our PD measure showed that periods of control were particularly informative about mass. Since the large majority of interventions were on one or other target puck, this suggests that most active manipulations of these objects were directly informative about mass. This is intuitive since a heavier object will be slower to accelerate for the same application of force (here, for the same displacement between the mouse and the controlled object center; cf. Figure 2b). While active participants were able to make use of this information, yoked participants' failed to capitalize on this. This suggests that an ability to anticipate the consequences of ones actions (i.e., where the mouse will be) may have been crucial to interpreting evidence stemming from the direct effects of control. Force information was higher on average during interventions but inspection of the individual-trial timelines shows that this was not always the case (cf. Figure 7). Periods of high force information were generally those where the target pucks were close to one another, far from other pucks and traveling slowly. Surprisingly, our information model considered the trials to contain stronger evidence about mass than about force while most participants found force easier to identify than mass. We return to this divergence in the General Discussion.

The quality of the control exerted by the active participants was an important determinant of the quality of the final evidence available to the yoked participants. However, the substantial drop-off from active to yoked accuracy is consistent with the idea that first-hand knowledge of *what* was being tested (e.g. force or mass), *when* and *how*, was likely to be crucial for learning successfully (cf. Markant & Gureckis, 2014a).

In Experiment 1, participants were asked about two properties of the worlds at the end of each learning trial. This means that there is ambiguity about what active learners focused on at any one moment. In Experiment 2, learners were asked to infer the value of a single property per trial. This makes it possible directly assess property-specific differences in active learning strategies. Furthermore, by having some yoked participants answer the same question as active participants but others answer a different question, we can assess directly how much it matters that learning goals are matched.

Experiment 2: Matched and mismatched yoking

In this experiment we had participants learn in the same worlds as before, but this time there were two blocks, one in which they were asked about the target objects' mass, and one in which they were asked about the force between the target objects. We then matched active participants with passively observing participants who were either asked about the same property (yoked–match) or asked about the alternative property (yoked–mismatch). In this setup, our information measures assess the extent to which active participants generated information tailored to the specific question they were asked about, and investigate how this affected yoked participants.

In line with Experiment 1, we hypothesized that active participants would generate information relevant to the property of the world they were asked about. To the extent that active learners did not also generate substantial information about the alternative property *en passant*, we expected this to result in lower accuracy for yoked–mismatch participants compared to yoked–match participants.

In our analyses of Experiment 2, we will further unpack participants' active learning strategies both qualitatively and quantitatively. First, we will contrast the informational statistics of the actions performed depending on the goal (mass vs. force identification). Second, we will have independent human coders assign labels to every one of the active participants actions based on the strategies proposed in Table 2. This will allow us to assess their overall prevalence as well as whether there was a clear separation of strategy-use depending on which property of the current world was under investigation.

Methods

Participants. One hundred and twenty participants were recruited from Amazon Mechanical Turk (76 male, M $\pm$ SD age 37.1 ± 11.8) using psiTurk (Gureckis et al., 2016). Following Experiment 1, we aimed initially to collect 20 participants per condition as in Experiment 1 but due to an error in our counterbalancing scheme, the first 20 active participants shared the same task order (i.e. they always answered a block of questions about mass before force). As a result, we doubled the cell size to retain a balance between block orders in the overall data set. Participants were paid at a rate of \$6 per hour, plus performance-related bonuses (\$$0.83 \pm 0.25$).

Conditions. Participants were assigned to one of three learning conditions:

1. **Active** ($N = 40$) Participants could grab the pucks and drag them around with the mouse, exactly as in Experiment 1 except they were aware which of the two candidate

questions they would be asked at the end.

2. **Yoked–match** ($N = 40$) Participants watched replays of the interactions of an active participant and had to answer the same question about each world as the active participant that they were observing
3. **Yoked–mismatch** ($N = 40$) This condition was like yoked–match except that participants had to answer the opposite question. If the active participant was asked about the masses of the target objects, the yoked–mismatch participant would be asked about the force relationship between the two targets. If the active participant was asked about the force, the yoked–mismatch participant would be asked about the mass.

The first 40 participants were assigned to the active learning condition, the subsequent 80 participants were randomly assigned to either the yoked–match, or yoked–mismatch conditions.

Worlds. Each participant watched or interacted with the same 9 microworlds from Experiment 1. However participants faced each world twice, once asked about the target masses, and once about the target forces.

Materials and Procedure. Given that only one property was probed per trial, each world was run for 30 s, rather than the 45 s in Experiment 1. In other respects, the Box2D physics simulator was set exactly as in Experiment 1.

Participants were told which question they will be asked prior to starting the interaction. Participants did not have to wait until the end of the trial to indicate their response. To ensure that active participants were motivated to interact efficiently with the worlds, the available bonus for each trial diminished from 10¢ for correctly answering at the very beginning of each trial to 5¢ for correctly answering by the end of the trial. If participants answered the question correctly they would receive whatever bonus remained at the moment of their final interaction with the response options. Even if answering early, participants would have to wait until the end of the 30s trial to continue, to ensure there was no incentive beyond the bonus payment to hurry. This procedure means that we can use an active participant's decision time as a marker delineating their actions in service of answering the question from any filler actions they might perform once they are already sure of the answer. As in Experiment 1, each question was paired with a confidence slider that was not tied to the bonuses.

Participants first completed instructions relevant to their condition (see `https://github.com/neilbramley/active_physics`), answered comprehension check questions, and then completed two blocks of 10 trials. In one block of trials they were asked about target masses, in the other they were asked about the target force relationship.

The order of the question blocks were counterbalanced between subjects. For each block, participants interacted with or watched one practice trial followed by all nine test worlds in random order, as in Experiment 1. As before, the practice trials were indistinguishable from the test trials and were excluded from analyses.

The yoked–match participants faced the same 20 trials as the active participants, while the yoked–mismatch participants were asked the alternative question on all trials. That is, if the active participant faced the force block first followed by the mass block, the yoked–mismatch participants would answer the mass question during the first block and the force question during the second block. For all conditions, the practice trials were worlds one in the force block and five in the mass block (cf. Table 1).

Unlike Experiment 1, where the letter labels were different for each trial, this experiment always used "A" and "B".[17] The distractor forces (between the other five pairs of objects than the target-target pair) and puck colors were again drawn at random for each trial for active participants and these were repeated for yoked–match and yoked–mismatch participants. The experiment took 20.6 ± 8.7 minutes.

Results

Following the structure of Experiment 1's results section, we first report judgment accuracy by condition and question before turning to our information measures.

Accuracy. Figure 8 shows overall performance by condition. We analyze the two mass and force accurately separately since, in this Experiment, they correspond to different trials. In the active, yoked–match, and yoked–mismatch conditions respectively, participants answered (M ± SD) 61±23%, 50±22%, and 46±16% mass questions correctly. There was a main effect of condition on accuracy on the mass question $F(2, 117) = 5.6, \eta_p^2 = 0.09, p = .003$ with fewer correct mass responses for both yoked–match $t(39) = 2.4, p < .001$ and yoked–mismatch $t(39) = 3.3, p = 0.001$ participants relative to active participants. In the active, yoked–match, and yoked–mismatch conditions respectively, participants answered 73±23%, 73±18%, and 57±20% force questions correctly. Thus, there was also a main effect of condition on accuracy on the force question $F(2, 117) = 8.0, \eta_p^2 = 0.12, p < .001$. There was no difference between the number of correct force responses for active and yoked-match participants $t(39) = -0.06, p = .95$, but a substantial drop from active to yoked-mismatch $t(39) = 3.4, p < .001$. Across the two questions, there was no main effect of the block order on accuracy nor any interaction with condition.

[17] There are not enough letters in the alphabet to give unique labels to all forty target objects across the twenty trials.

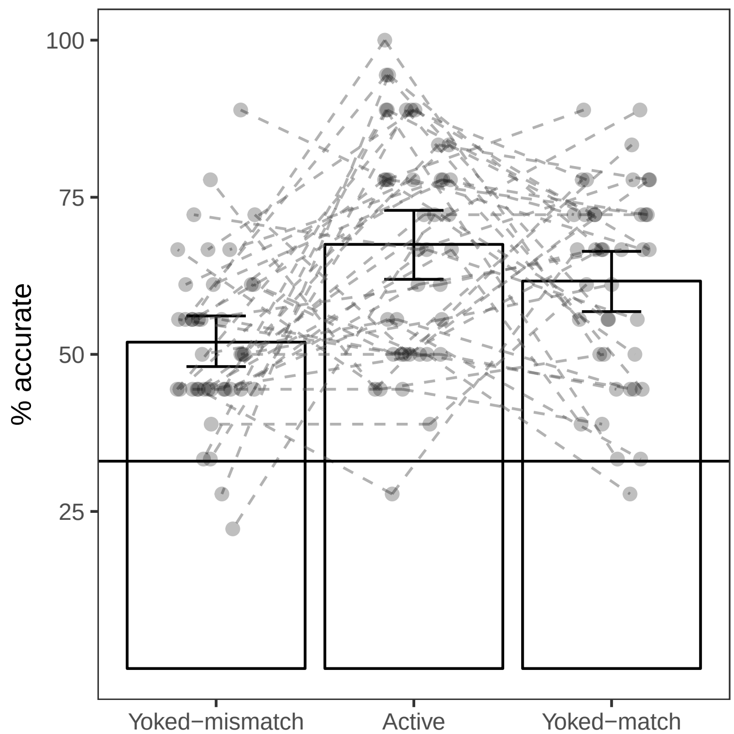


Figure 8. Experiment 2: Performance by condition measured by the accuracy with which participants' correctly answered the three-alternative forced-choice questions about the local forces and object masses. Bars denote condition means. Error bars denote bootstrapped 95% confidence intervals. Points denoting individual participants are jittered along the x-axis for visibility. Dashed lines connect active participants with matched yoked participants. The horizontal line indicates chance performance.

As in Experiment 1, participants were better at identifying the force relationship $67.9 \pm 21.8\%$ than mass $52.9 \pm 21.4\%$, with a repeated measures analysis revealing a substantial effect of question type on accuracy $F(1, 117) = 39, \eta^2_G = 0.11, p < .001$.[18] Twenty-six participants were more accurate on mass questions, 19 were equally accurate, and 76 were more accurate on force questions. A marginal interaction between question type and condition on accuracy $F(2, 117) = 2.6, \eta^2_G = .02, p = .075$, captures the slightly different patterns of accuracy for mass and force trials. Yoked–match participants were less accurate than active participants on mass trials $t(117) = -2.5, p = 0.014$ but not force trials $t(117) = -0.06, p = 0.95$.

As in Experiment 1, participants were more accurate at identifying when one of the targets was heavier than the other $63.5 \pm 19.3\%$ than when they had the same mass $54.2 \pm 22.2\%$ $F(1, 117) = 18.9, \eta^2_G = .05, p =< .001$ (see Figure 9). Similarly, participants

[18] η^2_G is a generalized measure of effect size recommended for repeated measures analysis (see, Bakeman, 2005).

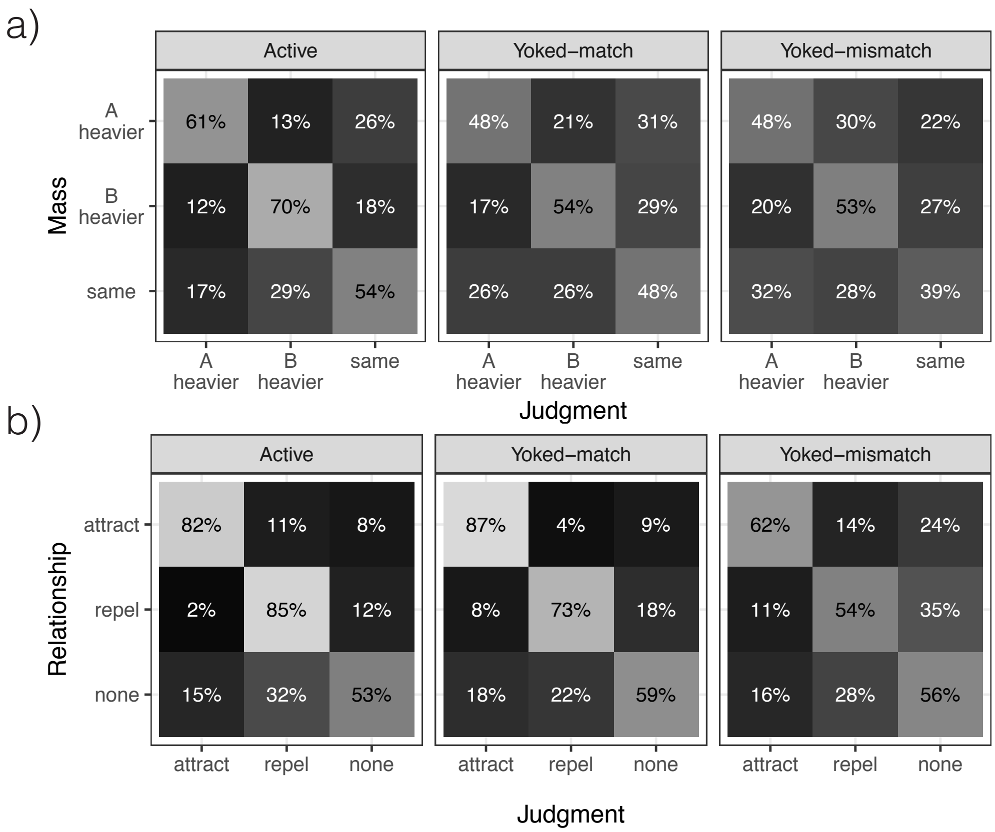


Figure 9. Experiment 2: Confusion matrices for mass trials (a) and force trials (b). For example, the mass question in the active condition, left matrix in a), when A was heavier than B (top row), participants correctly identified that this was the case 61% of the time, they falsely judged that B was heavier than A in 13% of the cases, and thought that both A and B have the same mass in 26% of the cases.

were better at identifying attraction $66.9 \pm 24.8\%$, followed by repulsion $59.6 \pm 23.7\%$, and worst at identifying when there was no force $54.6 \pm 24.1\%$, $F(2, 234) = 10, \eta^2_G = 0.05, p < .001$. In this experiment, by-answer accuracy patterns did not interact significantly with condition for either question type.

Response time. Participants made their judgments before the end of the trial 82.5% of the time, doing so after 18.3 ± 8.4 seconds on average. Response time did not differ significantly between the three conditions $F(2, 117) = 0.73, \eta^2 = 0.01, p = .48$, nor was it related to accuracy $r = 0.02, t(118) = 0.26, p = 0.79$.

Confidence. As well as being less accurate, participants were also less confident about their responses for mass $64 \pm 26\%$ compared to force $76 \pm 26\%$ questions $F(1, 117) = 62.4, \eta^2_G = 0.09, p < .001$. There was only a marginal difference by condition — active: $72 \pm 26\%$ yoked–match: $73 \pm 27\%$, yoked–mismatch: $65 \pm 26\%, F(2, 117) = 2.6, \eta^2_G = 0.03, p = 0.078$. Participants were around 11% more confident about their correct judgments $75 \pm 26\%$ than their incorrect judgments $64 \pm 27\%$ $t(118) = 7.9, p < .001$.

Measuring information. As in Experiment 1, we computed the posterior over models $P(W|\mathbf{d}, \beta; \mathbf{c})$ and associated posterior overall uncertainty $\mathrm{H}(W|\mathbf{d}, \beta; \mathbf{c})$, mass uncertainty, and force uncertainty with participants' judgments (see Figure 10). Here, we are interested in contrasting the evidence produced in trials where the active participant was asked about mass and where they were asked about force. A repeated measures ANOVA predicting posterior marginal uncertainty reveals a significant main effect of question type $F(1, 39) = 30, p < .001$, a very strong effect for property in question (mass or force) $F(1, 39) = 1252, p < .001$, and crucially, also a clear interaction between block and property $F(1, 39) = 56, p < .001$. This supports the hypothesis that active participants were successful at generating evidence that was informative specifically for the question they were asked to answer. This difference in information is behaviorally significant because it impacted performance, with commensurate performance by yoked–match to active participants but worse performance by yoked–mismatch participants.

Did the information differences correlate with performance? Average mass accuracy was inversely related to average mass uncertainty $t(39) = -2.6, r = -.39, p = 0.01$ for active participants, but not for their yoked–match counterparts ($p = .67$), nor for yoked–mismatch ($p = .14$) participants (on the relevant trials). Average force accuracy was strongly inversely related to average force uncertainty for active participants $t(39) = -6.0, r = -0.69, p < .001$ and their yoked–match counterparts $t(39) = -3.2, r = -.46, p = .002$ but there was no such relationship for yoked–mismatch participants on the relevant trials ($p = .35$). Thus, we see a clear relationship between lower uncertainty according to our model and higher accuracy for the active participants for both types of question. This carries over to yoked–match observers on the force question but not on the mass question.

Finally, at the individual trial level, the probability of answering a particular question correctly was significantly related to posterior uncertainty on the relevant dimension, even after accounting for the main effects of condition, dimension and the average accuracy for each trial type (i.e., those in Table A1). This is shown by a generalized logistic mixed effects model, with condition and question as covariates and random intercepts for each of the nine trial types, under which lower posterior uncertainty increases the probability of a correct response $\beta = -0.57 \pm 0.13, z = -4.5, p < .001$.[19]

[19]This was fit using R's `glmer` function in the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015).

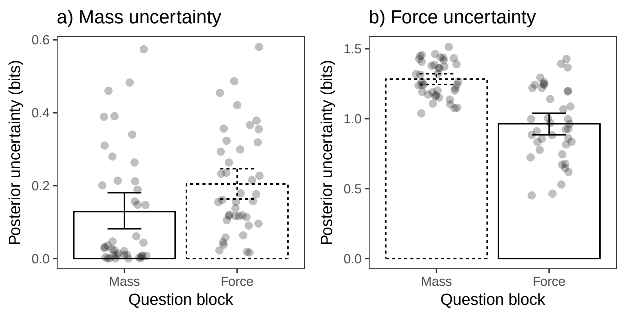


Figure 10. Experiment 2: Mean marginal posterior uncertainty about mass and force relationship according to IO model . Dashed bars correspond to the evidence available for yoked-mismatch participants. *Note*: Smaller bars indicate lower uncertainty (i.e., greater certainty). Error bars denote bootstrapped 95% confidence intervals. Points denoting individual participants are jittered along the x-axis for visibility.

Classification of participants' actions

To see whether the strategies identified in Experiment 1 were used differentially on mass and force questions, we asked two independent coders to label all 2313 interventions performed by active participants in Experiment 2 using the video coding software DataVyu (2014). Coders were blind to the learners' goals and our hypotheses. We used the 7 categories described in Table 2: (a) Launching, (b) Knocking, (c) Throwing, (d) Shaking, (e) Encroaching, (f) Deconfounding and (g) Controlling. In addition, we included the following two categories:

(h) **Multiple**. The intervention satisfies more than one of a–g and none of the strategies is clearly dominant.

(i) **Unclear/other**. The intervention does not clearly fall under any of a–h.

Coders were provided with the descriptions and diagrams as in Table 2, and a short video of an intervention exemplifying each strategy taken from Experiment 1. If a control period appeared to fall under multiple categories but one of these categories was clearly dominant, the coder was instructed to select that category rather than 'Multiple'.[20] Inter-rater agreement on the primary category was .79, and Cohen's $\kappa = .76 \pm 0.02$, both higher than their respective heuristic criteria for adequacy of 0.7 and 0.6 (Krippendorff, 2012; Landis & Koch, 1977). The agreed codes are summarized in Figure 11 and a confusion matrix showing where coders disagreed is included in the Appendix. The only clear difference between the two coders is that the second coder was more likely to select 'Unclear' or 'Multiple'.

[20]All materials given to coders along with coded videos are available in the online repository.

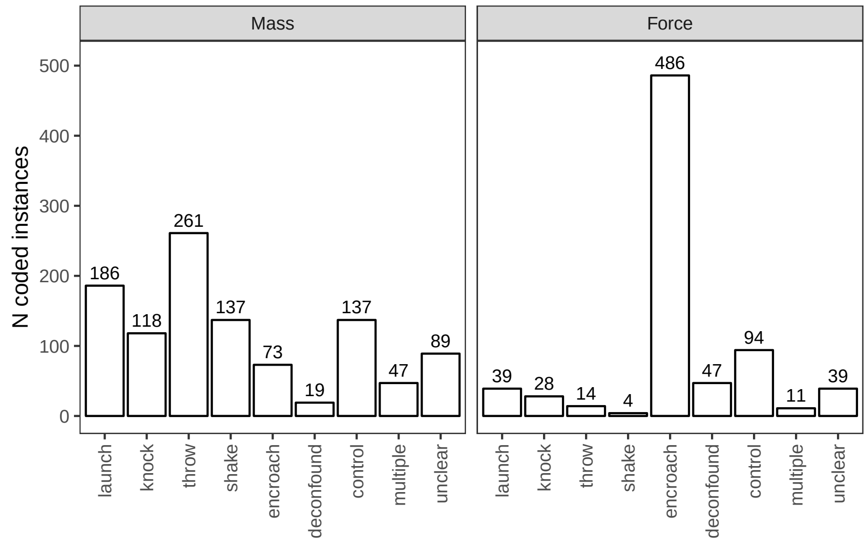


Figure 11. Coded strategy use in Experiment 2. Bars show the N agreed instances of each of the strategies from Table 2 broken down by question (panels).

We found large differences in the distribution of strategies by condition $\chi^2(8) = 838, p < .001$. When focused on mass, active learners perform many 'Launches', 'Knocks', 'Throws' and 'Shakes'. When focused on the force relation they rarely perform these sorts of actions, and instead predominantly 'Encroach' (move and holding targets close together).

Predictive divergence

As in Experiment 1, we computed the online information measures PD_{mass}, PD_{force} and baseline PD throughout every trial for each participant.[21] Recall, PD lets us look inside the trials to assess whether the mass and force evidence was produced predominantly during periods of control or during periods of passive observation. Furthermore, we can explore the extent to which interventions ostensibly targeted at revealing one property also reveal the other property.

Figure 12 compares mean PD scores during periods of observation and periods of intervention as in Figure 6b. Baseline is lower during periods of intervention than during periods of observation for both mass $t(35) = -2.5, p = 0.016$ and force $t(35) = -6.6, p < .01$ question blocks.[22], and is generally considerably lower than both

[21]We provide predictive divergence timelines, as in Figure 7a, paired with movie replays for all participants in the online repository.
[22]Some participants in the active condition did not intervene at all either across all mass trials, across all force question trials or both. Hence, these participants were not included in this analysis.

$\text{PD}_{\text{mass}}\, t(39) = -16.8, p < .001$ and $\text{PD}_{\text{force}}\, t(39) = -11.4, p < .001$. PD_{mass} is considerably higher during the mass question block 0.206 ± 0.044 than the force question block 0.16 ± 0.023 $t(39) = 6.8, p < .001$, and over periods of intervention compared to observation on both mass question trials $t(35) = 12.6, p < .001$ and force question trials $t(35) = 11.2 p < .001$. $\text{PD}_{\text{force}}$is considerably higher during the force question block 0.20 ± 0.053 than the mass question block 0.12 ± 0.028 $t(39) = 8.7, p < .001$, but exhibits a different control pattern with marginally lower PD_{force} during interventions (compared to observation) on force question trials $t(35) = -2.24, p = 0.031$ but substantially lower PD_{force} during interventions on mass question trials $t(35) = -5.5, p < 0.001$.

Overall, this pattern confirms what was suggested by the results of Experiment 1. Simply taking control of a target object provides evidence about its mass. The heavier object reacts more sluggishly than the lighter object. Thus, PD_{mass} was raised when interacting with the objects regardless of the goal. In contrast, PD_{force} was spread over both periods of control and subsequent periods — e.g., while the pucks remained in close proximity — but was only raised overall on trials where active learners' goal was to identify force. The lower baseline PD during interventions suggests that participants were also somewhat successful in minimizing the confounding influence of the surrounding dynamics.

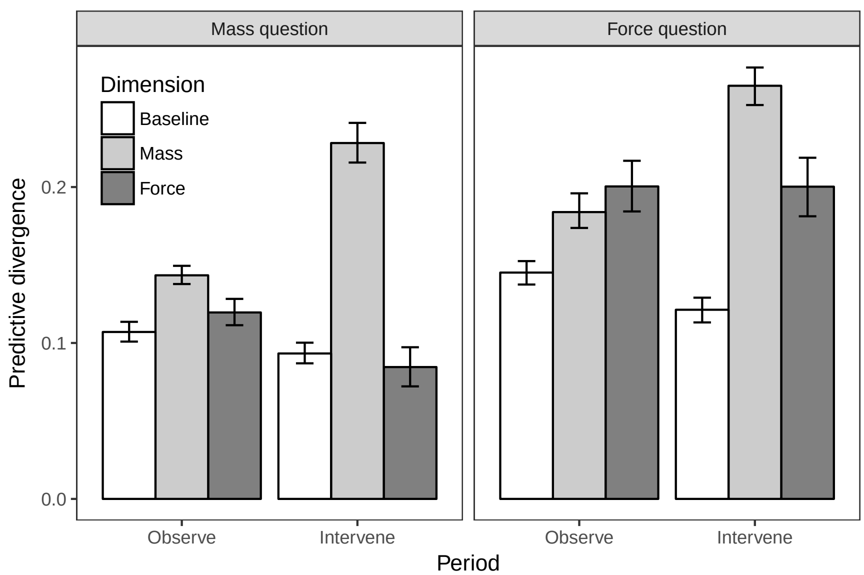


Figure 12. Experiment 2: Comparison of average predictive divergence for periods of observation and intervention (i.e., controlling one of the pucks). Larger values indicate more evidence. Bars denote means and error bars denote bootstrapped 95% confidence intervals. "Baseline" denotes the average PD measured for every dimension over which the worlds in $\mathcal{W}$ vary.

We now take a closer look at the PD measures by relating them to the different coded intervention strategies (cf. Table 2 and Figure 11). To do this, we took instances of each strategy for which both coders agreed, timelocked them to their onset, and averaged our PD measures over a short period period beginning 1s before each onset. The resulting profiles are in the righthand column of Table 2. By plotting these profiles along with the distribution of offset times, we see that ‘Knocking’, ‘Throwing’ and ‘Shaking’ are all relatively punctate actions, lasting around a second. Their profiles also confirm our intuitions about the nature of the evidence they produce. ‘Launching’ — throwing a puck at a target puck — causes a spike in PD_{mass} shortly *after* the controlled object is released i.e., around when the pucks collide. ‘Knocking’ — hitting a target puck with a still-controlled puck — reaches its maximum before release (again presumably when controlled objects collide), while ‘Throwing’ — avoiding the other pucks — creates a smaller spike since only one controlled object is involved, and this also occurs before release. ‘Launching’ and ‘Knocking’, but not ‘Throwing’, are associated with a small spike in PD_{force}, presumably because they often involve targets passing close to one another. ‘Shaking’ also leads to strong and sustained increase in PD_{mass}. Instances of ‘Shaking’ are variable in length and do not create a spike in PD_{force}, since during shaking other objects must be avoided. Encroaching actions are also long and variable , but lead to the substantial and sustained increase in PD_{force} continuing after release. ‘Deconfounding’ — moving non targets out of the way — is also associated with a sustained period in which PD_{force} is higher than PD_{mass} (although lower than for encroaching). By reducing probability of collisions, ‘Deconfounding’ leads to situations with little evidence about the target objects’ mass.

We can also ask which strategies were associated with successful identification of the target properties. At the participant level, we entered the frequency with which a participant performed each of the nine coded actions as independent predictors of accuracy on the two question types. In both cases, two of the nine strategies were significantly associated with performance and in combination they explained 58% of the variance in accuracy on both questions. For mass trials, ‘Shaking’ was most strongly associated with accuracy $\beta = 0.022, t(30) = 3.9, p < .001$, followed by ‘Throwing’ $\beta = 0.016, t(30) = 3.3, p = .003$ while the other seven strategy codes did not have a significant association. For force trials, ‘Encroaching’ was strongly related to accuracy $\beta = 0.016, t(30) = 4.2, p < .001$ while ‘Launching’ was significantly negatively related $\beta = -0.034, t(30) = -2.2, p = .04$. Again, the other 7 strategy labels did not have a significant association with accuracy. There was no evidence that participants adapted their use of strategies systematically over the 9 test trials, with no relationship between the number of strategies applied and trial position within either mass $p = .81$ or force $p = .17$

question blocks. However, we note that each block was preceded by a practice trial which we did not analyze, so it is possible that some adaptation took place between the practice and test trials.

Finally, we can ask whether the goal-specific value of an action (i.e., its average predictive divergence) predicts participants' tendency to perform that action. To do this, we calculated a single predictive divergence value for each strategy and goal. For example, we calculated how useful the 'Encroaching' or 'Shaking' strategy are for the goal of identifying the local force and the objects' mass. We took the mean divergence on the relevant dimension over the period from the onset of a learner's control to one second after they released control (to allow for a delay in the effect of some of the actions such as 'Launching'). We then averaged these across all coded instances of each strategy. The resulting values predict the frequency with which participants performed the different strategies across the force and mass trials $r = .75, t(12) = 3.8, p = 0.02$.

Discussion

In Experiment 2, we found that active learners interacted with the microworlds in ways that served to emphasize relevant physical properties for their current learning goal and mitigate confounding information. Observers whose goal it was to identify the same property as the active learners were as accurate on average as their active counterpart, while those whose goal it was to identify the other property were considerably less accurate. This supports the idea that ambiguity about the active learner's goal in Experiment 1, may have driven the reduction in accuracy for yoked observers.

By having human coders classify all of participants' active learning actions, we related these differences in overall accuracy to differences in strategy selection. When focused on mass, participants repeatedly threw the target objects at one another or into empty space, knocked them together, and shook them. Active interactions involving the target pucks tended to be informative about mass even if this was not the learners' goal, but this evidence quickly dissipated. When learning about forces, participants used most of their control actions to bring the target pucks close together ('Encroaching'), creating unambiguous and extended demonstrations of the target relationship that often continued beyond the point at which the pucks were released. In both trial types participants took actions (like 'Controlling' and 'Deconfounding') that reduced the complexity of these inferences by reducing how dependent the outcomes of the tests were on the distractor and non-target properties. This is shown by the reduced Baseline PD during interventions in Figure 12.

The most effective strategy for identifying masses in terms of the size of its PD spike and association with performance was 'Shaking', while 'Encroaching' played a dominant role for force trials. Overall, the different strategies' average predictive divergence was positively related to participants' propensity to select that action for a given goal. Interestingly, the quality of active learners' force-revealing actions predicted yoked learners' accuracy but the same was not true for mass, suggesting that first-hand experience and anticipation of actions while controlling the pucks was important for making use of the extra evidence generated by those interventions that are diagnostic of mass.

General Discussion

Intuitive examination of our everyday experience suggests that we move and interact with the physical world to reveal latent properties. But how effective are those actions at helping us meet our learning goals? This is an interesting cognitive problem exactly because it is so under-constrained. Here we take the first step toward understanding this behavior by analyzing quantitatively how strategies reveal information in real time about latent properties in a virtual physical world. We developed a framework for assessing the evidential value of interactions with a simulated physical world applied it in two experiments in which participants actively learned about the properties of objects in simple 2D physical "microworlds". Experiment 1 revealed that participants who were able to grab and manipulate the objects, are more accurate at identifying latent object properties than passive or yoked observers. Our analysis revealed that part of this advantage stems from the fact that active participants actions produced stronger evidence about the target properties than was produced by the naturally occurring dynamics. In Experiment 2, we investigated whether the worse performance by yoked participants was due to uncertainty about the active participant's current goal. Yoked participants who were given the same learning goal no longer underperformed their active counterparts overall (but were still somewhat worse on mass judgments), while those with a different learning goal remained disadvantaged. Going beyond our primary aims, we also found systematic differences in the strategies people adopted dependent on their learning goal that we classified (through a combination of video coding and information measurement) as a set of micro-experimental strategies. In summary, our key findings are:

1. Human intuitive experimentation in physical environments depends systematically on learning goals and is formally effective at revealing physical properties of interest.
2. Learning from observing another person's actions is only successful when the goal is shared and unambiguous, and the evidential signal does not depend on an ability to anticipate the other person's next action.

3. Learners' sequences of actions are characterizable as micro-experiments that emphasize particular properties while minimizing the confounding influence of uncertainty about other properties.

We now discuss these findings more broadly in the light of ongoing and related work. We first discuss our simulation framework, then revisit differences between active and yoked observational learners before turning to the questions of how learning strategies are themselves learned and the connections between naturalistic active learning and control.

Simulation

The idea that we use mental simulations to reason about physics raises important questions about how such a simulator could be implemented in the brain, as well as questions about its specificity, optimality and ability to deal with uncertainty (Davis & Marcus, 2016; Davis, Marcus, & Chen, 2013; Davis et al., 2013; Marcus & Davis, 2013). While much of the details of what form this intuitive understanding takes still need to be worked out, the results of these and number of recent experiments are consistent with the view that people have a rich intuitive theory of physics that supports approximately accurate mental simulations of several aspects of physical scenes (Gerstenberg et al., 2012, 2015; Smith et al., 2017). Thus, although deeply related, in the current paper we remained agnostic about *how* people represent and reason about the physical world and instead focused on the question of whether actions human learners take in the dynamic physical worlds are informative about latent physical properties. In particular, we viewed this question from the perspective of an Ideal-Observer analysis (Geisler, 1989; Kersten, Mamassian, & Yuille, 2004; Marr, 1982). This allowed us to quantify the information available to the learner, and the amount of information generated by the learner's actions, without making specific claims about the psychological mechanisms that underlies people's understanding of physics. To be clear, our intention is not to claim that people are near-normative at learning in this domain. The normative model provides a useful benchmark to compare participants' actions and judgments against.

Information measures and detectable differences

One difference between participants' judgments and our information measures was that the models suggested there was more information available about mass than force, while around two thirds of participants still found the mass question more difficult to answer than the force question. There are several possible explanations for this. One is that participants were also uncertain about other aspects of the worlds which were not included in the simulations. For simplicity, our ideal observer model started its simulations with the true locations and velocities and had accurate knowledge of the worlds' fixed properties (e.g., the elasticity of the objects, the friction, the strength of the attractive force of the mouse on controlled objects, the laws of the simulated physics). It could be that incorporating uncertainty about these other aspects makes the model likelihoods and predicted divergences less sensitive to differences in mass.

Another possibility is that participants may have made inferences that go beyond what our ideal observer model captures. For instance, participants experienced attraction and repulsion among the non-target as well as between the target objects. Thus, participants may have learned about the characteristic behavior of attraction and repulsion partly by comparing against these other objects' motion. Our ideal observer model does not capture this between-object generalization of expectations about dynamics, but could be extended to do so by including shared, or hierarchically related parameters (Kemp, Goodman, & Tenenbaum, 2010; Lake, Ullman, Tenenbaum, & Gershman, 2017).

A third possibility is that the kinds of divergences caused by the local forces are more easily spotted by our perceptual system. The local forces created *qualitative* differences in the paths of objects (e.g. making objects veer toward or away from one another rather than continuing in a straight line) while the masses affected things more *quantitatively* (e.g. affecting the degree of veering or the angle of exit from collisions). Our ideal observer model assessed likelihoods using the distance between simulated objects' projected motion in terms of magnitude r and angle θ. Separating these aspects of the likelihood (as in Figures A2 and A3 in the Appendix) reveals that the mass–force evidence asymmetry is present in the magnitude information but not the angle θ information. It was not possible to fit parameters of a multivariate likelihood model due to the large size of the data, so we opted to scale r and θ by their empirical variances so that they contributed roughly equally to our model predictions. However, it is plausible that the perceptual system is better tuned to detecting change in direction than change in velocity (Tenenbaum & Witkin, 1983; Treisman & Gormican, 1988), especially as veering movements are hallmarks of causal influences (Michotte, 1946/1963; Oakes & Cohen, 1990; White, 1995).

In general, the use of a quantified "distance" between a simulated and an observed outcome in place of a proper likelihood function, is a new and growing area of applied machine learning sometimes called "likelihood-free inference" (Gutmann, Corander, et al., 2016). For example, finding parameter settings that minimize distance between forward simulations and observations has proved an effective way to learn complex generative models in particle physics (Brehmer, Freitas, López-Val, & Plehn, 2016) and systems biology (Ratmann et al., 2007). However, a general issue with this approach is identifying a good distance measure. As such, it is instructive to examine what choices of distance measure can explain the behavior of nature's successful learners. Our assessment of a range of distances for driving our IO model in the Appendix supports the idea that humans rely on motion information, particularly direction change for physical inference.

Diagnosing differences between yoked and active learning

There are a number of ways one might model the differences between active and yoked performers' experiences. In our analysis, we treated all objects' locations and velocities as equally uncertain. However, it is plausible that active learners have a better idea about the locations of objects while controlling them since they can incorporate direct motor feedback from their mouse or finger on the track-pad (e.g. Körding & Wolpert, 2004). One could model this by assuming smaller perceptual uncertainty for objects under control (captured by the β parameter in our model). This would result in the prediction that active learners receive stronger evidence from events involving the controlled object. Additionally, learners' attention is certainly limited relative to the action in the scenes. We cannot easily track the location of multiple objects at the same time (Scholl, 2001; Vul et al., 2009). Thus, we might model learners' attention as a focal window. Active learners could then use their knowledge of planned action to focus their attentional window on a region they expect to be informative. Yoked learners lack this foresight and hence are more likely to be attending elsewhere when something informative happens.

A more mundane reason for why yoked learners might perform differently to active learners is a difference in motivation and engagement with the task. Interactivity is often seen as increasing engagement, leading to improved learning irrespective of differences in information or a better match between evidence and processing (Berlyne, 1960; Gureckis & Markant, 2012; Hebb, 1955). However, we attempted to equate motivation across conditions by incentivizing participants. Furthermore, in Experiment 2 yoked–matched participants performed as well overall as active participants while yoked–mismatched participants were worse. This suggests that, when an active learner's goal is unambiguous and aligned to an observer's, the observer is able to make commensurate use of the evidence.

Learning to actively learn

Computing ideal interventions requires playing out all possible possible outcomes of all possible actions in all possible worlds (Raiffa, 1974). This calculation is intractable for any non-trivial world and plausibly-bounded learner. Fortunately, many active learning situations share fundamental similarities. We are repeatedly faced with uncertainty about specific aspects of our current environment: How heavy is that suitcase? Can this table support my weight? Will the surfboards stay on the roof of the car? Fortunately, the laws of physics, as well as many of the materials involved, are shared across these everyday contexts. Once we have a sense for what actions allow us to resolve uncertainty about a familiar physical property, this can provide a strong prior about what will be effective in the future, analogous to the way domain expertise supports one-shot passive inference (Goodman et al., 2011; Vul, Goodman, Griffiths, & Tenenbaum, 2014). This perspective may explain why humans are robustly successful at gathering information even though calculating expected information is intractable. While transferring strategies from past experience helps directly, having a good intuitive theory of physics, and of our own actions in the world makes us more adaptable.

Several recent papers in machine learning have begun to explore learning to actively learn in dynamic environments. For example, Denil et al. (2017) use reinforcement learning to train deep neural networks to perform simple actions with the goal of identifying the heaviest of a set of virtual blocks, or the number of (potentially occluded) blocks in a tower (see also Chang, Ullman, Torralba, & Tenenbaum, 2016). Relatedly, Agrawal, Nair, Abbeel, Malik, and Levine (2016) train a robotic arm to actively poke and prod real physical objects and successfully predict their resultant displacement. These projects found success through explicitly or implicitly encoding an abstract model of the action and state spaces. This allows planning to be done in a top down way, one step removed from the detailed realization of actions and dynamics (i.e. in pixel space). In the current task this is like having a model that supports planning at the level of strategy selection, that is, whether to 'Shake' or 'Encroach' given your inquiry goal.

Some developmental psychologists have argued that children are "intuitive scientists", with inquiry skills that are either innate or established early during development (Gopnik, 2012; Gopnik et al., 2004; Gopnik & Sobel, 2000; Lucas, Bridgers, Griffiths, & Gopnik, 2014). However, this perspective has not yet fleshed out how these scientific intuitions are learned and applied across contexts, at least not to the extent that "intuitive theories" have been in explaining few- and one-shot judgments (e.g. Gerstenberg & Tenenbaum, 2017; Lake et al., 2017; Tenenbaum, Griffiths, & Niyogi, 2007). We see the current experiments as supporting the idea that adults are intuitive physical scientists in this sense. Our participants were able to draw on their intuitive understanding of physical dynamics when interacting with simulated physical objects, repeatedly applying strategies that presumably work well in the real world, or that they discovered to be effective in earlier trials. A nice example of such cross-trial transfer was the case of 'Shaking'. We did not come up with 'Shaking' when designing and piloting the task, but 21 out of 40 participants in Experiment 2 discovered this strategy and used it multiple times in their remaining mass block trials. Our PD analysis (e.g. Table 2) suggests that 'Shaking' is a particularly effective way of answering the mass question: participants who used this strategy answered the mass questions more accurately.

Notwithstanding their systematicity, the strategies we observed still largely lack some key properties of formal experiments that help rule out artifacts and experimenter bias (Winer, 1962). Our learners could not perform multiple actions in parallel, meaning that there is generally no direct "control group" to compare an intervention against, and it seems unlikely that random allocation played much role in participants' choices of actions. However this is a fundamental problem for online naturalistic learning, and presumably one we must find ways to sidestep where possible.

Participants in our tasks certainly repeated actions many times, in this way potentially averaging over variation in surrounding conditions. Furthermore, in line with the mechanism of our inference models, they may often have made property inferences by synthesizing an appropriate comparison through memory and simulation. Gerstenberg et al. (2017) use eye tracking to show that when judging the causal role of a collision event, participants' eyes "play out" the counterfactual trajectory of objects if the collision had not occurred. This suggests people estimate the causal effect of everyday physical interactions by comparing what actually happened against expectations about what would have happened otherwise. In a similar way, to learn in the current task, people must often have been able to make informal comparisons between observations and either relevant memories of past dynamics or synthesized expectations of dynamics. In the current context for example, the learner cannot shake both objects at once so must shake A and then shake B and compare the current motion against the recent memory. The behavior of two objects during encroaching may be compared against the behavior in other trials, or for other pairs of objects.

An interesting question is when, during development, children start to exhibit the systematic and goal-directed active learning we observed in adults. Some studies have suggested that children's exploratory behavior is essentially random (Cole, Robinson, & Adolph, 2016; Kretch & Adolph, 2017) while others have emphasized ways in which even young children are sometimes efficient active learners (Bonawitz et al., 2011; McCormack, Bramley, Frosch, Patrick, & Lagnado, 2016; Schulz & Bonawitz, 2007). The former research has typically focused on children's behavior in "real world" situations — for instance, analyzing whether children walk purposefully around a laboratory playroom. The latter has focused on more constrained tasks that lend themselves straightforwardly to formal analyses — in which children select from a limited set of possible actions. One interesting exception is Stahl and Feigenson (2015), who show 11 month olds exposed scenes that violate "core" (Spelke & Kinzler, 2007) principles, act in different ways depending on which principle is violated. After an object appears to pass through a wall in apparent violation of solidity, the infants tended to "bang" it, but having seen it appear to roll off an edge without dropping, were more likely try to "drop" the object. This is suggestive that rudimentary strategies for active physical inference may emerge very early. However, the rudimentary classification precludes a more detailed assessment. Thus, a variant of this task, perhaps using a touch screen interface, will be valuable for this debate. We can allow children to engage in naturalistic, albeit simulated, real-time active learning, record behavior in fine detail, and use our information framework to quantitatively assess the systematicity of their actions.

Heuristics

Ullman et al. (2014, to appear) proposed a heuristic model of participants' passive learning in their task. The basic idea was that learners collect statistics about a world they are observing — for example, the pucks' average positions, velocities and pairwise distances — and compare these to summary statistics of past observations (or internal simulations) of worlds with known properties. For example, objects that attract one another tend to be closer together, so when two objects are close together on average in a scene, the heuristic will assign a high probability to an attractive force. This approach is more frugal than the ideal observer model because the statistics do not depend on online simulation. However, this approach is not directly applicable to the active setting. For instance, if you repeatedly move pucks close together, then their average location is no longer a good guide to their force relationship. However, a similar idea could be powerful within the expanded framework we outlined above. For example, relative to an established active learning strategy it could be enough to identify abstract statistical properties of the outcomes. Given equal 'Shaking', or 'Throwing' of two objects of potentially different masses, will the heavier one move *more* or *less* than the other? Given 'Encroaching', will an attractive object veer *toward* or *away*? Such heuristic qualitative decision criteria could be learned through experience and preplay (Pfeiffer & Foster, 2013). Once the characteristic effects of interventions have been learned, an expert active learner need merely look out for some critical signal in the dynamics following their intervention, without having to perform costly online simulation and comparison.

The road to control

We first introduced interventions as behaviors that affect the world without being caused themselves (Pearl, 2000). However, the continuous and interactive nature of the current task challenges this distinction. Active learners are likely to have planned their actions based on their current beliefs and learning goal, tried to perform the action as planned, and then updated their beliefs. This would be a reasonable idealization of some of the more punctate actions such as 'Throwing' and 'Launching'. However, many of participants' actions were extended in time, and clearly reactive to the ongoing dynamics. For instance a learner might shake an object harder — moving the cursor faster and further — if it reacts sluggishly. When 'Encroaching', a learner might end up "chasing" or "dragging" one target with another depending on how it reacts, and in ways that were not plausibly predictable in advance. Intuitively, these reactive behaviors themselves become an important part of the evidence produced by the interaction. For instance, mass could be assessed by comparing how hard an object must be shaken (how far and fast to move the mouse in this task) to make it oscillate a certain amount. It would be just as reasonable to infer mass this ways as to judge how far it oscillates for a fixed amount of 'Shaking'. For 'Encroaching', having to constantly chase the other target is a clear mark of repulsion, while being able to drag the other object is a mark of attraction. These examples are perhaps rather idiosyncratic to the current task. In the real world, we intuitively assess how heavy objects are by picking them up and monitoring how much effort we had to exert to do so. In these cases, the primary evidential signal seems to come from the reactive control required to achieve a goal, rather than from how the world reacts to a fully preplanned action.

The above considerations suggest that, as the learning context becomes more interactive and open-ended, active learning becomes increasingly closely related to an adaptive , or "dual-", control problem (Feldbaum, 1960; Guez, 2015; Klenske & Hennig, 2016; Schulz, Klenske, Bramley, & Speekenbrink, 2017). The idea is that in many real world contexts, we face the "dual" problem of learning how something works while already "on the job". As a simple example, learning how to play tennis takes place, largely, while attempting to play tennis. At first our shots are wild and do not go where we intend. But, over many games, we learn to adapt our swing to different angles and speeds of the incoming ball. Eventually we may become experts with a sophisticated control model of tennis that allows us to get the ball where we want most of the time. Analogously, participants in our task might have learned, in part, by attempting to move objects to particular locations and adapting their control model of the world in the process, before probing it to answer the test questions. In general, as we study richer, more physically embedded cognition, it seems likely that we will see an increasing convergence between models of active learning and adaptive control.

The real world

The simulation-based framework we used in our task and analyses allows us to precisely record participants' real-time interactions and analyze them within a computational framework that links idealized probabilistic inference with information-driven action selection. We noted at the start that the microworlds we explored here are a simplification of true physics. However, the methods we used to study participants' behavior in these worlds are very general and can be extended to more physically realistic scenarios. In general, with this approach we can start to model interactive behavior in any simulable scenario, whether information seeking or reward driven.

Our inference framework is a form of analysis by synthesis where we assume the learner strives to build an internal model that can produce the dynamics they observe (Yuille & Kersten, 2006). The novelty here is that analysis is facilitated by having a precise record of a participants' experience. Furthermore, participants probative actions provide a useful source of evidence about their learning trajectory, promising insights similar to how eye-tracking is used to model attention and simulation (Gerstenberg et al., 2017).

Using our framework, computer and mobile game data could be used as case studies of human adaptive control (cf. Mnih et al., 2013). Furthermore, the advent of augmented reality displays will make it possible to study action in the physical world directly, precisely recording and partially controlling an agent's interactions with a real environment. Using virtual reality devices like force-feedback gloves will also allow us to overcome the largest remaining discrepancy between the nature of the experience of the task and that of real world learning; the feeling of overcoming static forces, such as holding an object against gravity or two magnets together against repulsion. In general, we feel it is crucial to closely study behavior in realistic tasks. Such tasks better capture the immediate and dynamic nature of real-world evidence, and resemble more closely the continuous learning and control problem that characterizes the human experience.

Conclusions

Making sense of the physical world is one of the most fundamental problems for cognition. The physical world is the primary source of data, as well as the ultimate location of rewards. We are also intimately connected to it, affecting it in one way or another with every action we take. Despite this, much of the research on the computational basis of active learning has stayed far from the coal face, typically studying learning from abstract (normally passive) presentations of information that lack meaningful physical extension. This is understandable since the complexity of natural environments complicate the application of mathematical tools such as probability and information theory. However, these tasks underestimate the richness and immediacy of the data that learners have access to. The studies and analysis we present here take an important step toward bridging this gap. Our task dynamics have a richness and immediacy familiar from the real world, yet the simulation is still constrained enough to submit the results to familiar formal methods used in analysis of prior active learning tasks. Most striking was the sophistication of the active control strategies that participants used to reveal specific properties. How these strategies are learned and applied in development is an interesting and open question in psychology, and a formidable challenge in the search for more human-like artificial intelligence.

References

Agrawal, P., Nair, A., Abbeel, P., Malik, J., & Levine, S. (2016). Learning to poke by poking: Experiential learning of intuitive physics. *arXiv preprint arXiv:1606.07419*.

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior research methods*, *37*(3), 379–384.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, *110*(45), 18327–18332.

Berghman, L., Goossens, D., & Leus, R. (2009). Efficient solutions for mastermind using genetic algorithms. *Computers & operations research*, *36*(6), 1880–1885.

Berlyne, D. E. (1960). Conflict, arousal, and curiosity.

Best, J. B. (1990). Knowledge acquisition and strategic action in "mastermind" problems. *Memory & Cognition*, *18*(1), 54–64.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

Bottema, O., & Roth, B. (1979). Theoretical kinematics. *NORTH-HOLLAND PUBL. CO., N. Y., 1979, 558*.

Bramley, N. R., Dayan, P., Griffiths, T. L., & Lagnado, D. A. (2017). Formalizing Neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, *124*(3), 301–338.

Bramley, N. R., Gerstenberg, T., & Tenenbaum, J. B. (2016). Natural science: Active learning in dynamic physical microworlds. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society* (pp. 2567–2573). Austin, TX: Cognitive Science Society.

Bramley, N. R., Lagnado, D. A., & Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through interventions. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *41*(3), 708–731.

Bramley, N. R., Mayrhofer, R., Gerstenberg, T., & Lagnado, D. A. (2017). Causal learning from interventions and dynamics in continuous time. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Brehmer, J., Freitas, A., López-Val, D., & Plehn, T. (2016). Pushing higgs effective theory to its limits. *Physical Review D*, *93*(7), 075014.

Broadbent, D. E., FitzGerald, P., & Broadbent, M. H. (1986). Implicit and explicit knowledge in the control of complex systems. *British journal of Psychology*, *77*(1), 33–50.

Bruns, H. (1887). Über die integrale des vielkörper-problems. *Acta Mathematica*, *11*(1), 25–96.

Chang, M. B., Ullman, T., Torralba, A., & Tenenbaum, J. B. (2016). A compositional object-based approach to learning physical dynamics. *arXiv preprint arXiv:1612.00341*.

Chater, N., & Oaksford, M. (2013). Programs as causal models: Speculations on mental programs and mental representation. *Cognitive Science*, *37*(6), 1171–1191.

Cleveland, W. S., Grosse, E., & Shyu, W. M. (1992). Local regression models. *Statistical models in S*, *2*, 309–376.

Coenen, A., Bramley, N. R., Ruggeri, A., & Gureckis, M., Todd. (2017). Beliefs about sparsity affect causal experimentation. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Coenen, A., Nelson, J., & Gureckis, M., Todd. (submitted). Asking the right questions about human inquiry.

Coenen, A., Rehder, R., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Cole, W. G., Robinson, S. R., & Adolph, K. E. (2016). Bouts of steps: the organization of infant exploration. *Developmental psychobiology*, *58*(3), 341–354.

Datavyu Team, T. (2014). *Datavyu: A video coding tool.* http://datavyu.org. Databrary Project, New York University.

Davis, E., & Marcus, G. (2016). The scope and limits of simulation in automated reasoning. *Artificial Intelligence*, *233*, 60–72.

Davis, E., Marcus, G., & Chen, A. (2013). Reasoning from radically incomplete information: The case of containers. In *Proceedings of the Second Annual Conference on Advances in Cognitive Systems* (Vol. 273, p. 288).

Denil, M., Agrawal, P., Kulkarni, T. D., Erez, T., Battaglia, P., & de Freitas, N. (2017). Learning to perform physics experiments via deep reinforcement learning. *International Conference on Learning Representations*.

Feldbaum, A. (1960). Dual control theory. *Avtomatika i Telemekhanika*.

Fragkiadaki, K., Agrawal, P., Levine, S., & Malik, J. (2015). Learning visual predictive models of physics for playing billiards. *arXiv preprint arXiv:1511.07404*.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267.

Gerstenberg, T., Goodman, N., Lagnado, D. A., & Tenenbaum, J. B. (2012). Noisy Newtons: Unifying process and dependency accounts of causal attribution. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.*

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2015). How, whether, why: Causal judgments as counterfactual contrasts. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 782–787). Austin, TX: Cognitive Science Society.

Gerstenberg, T., Peterson, M. F., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2017). Eye-tracking causality. *Psychological science*, *28*(12), 1731–1744.

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In M. Waldmannn (Ed.), *Oxford handbook of causal reasoning* (pp. 515–548). Oxford University Press.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. *Psychological Review*, *118*(1), 110–9.

Goodrich, M. T. (2009). On the algorithmic complexity of the mastermind game with black-peg results. *Information Processing Letters*, *109*(13), 675–678.

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*(6102), 1623–1627.

Gopnik, A., Glymour, C., Sobel, D., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, *111*, 1–31.

Gopnik, A., & Sobel, D. M. (2000). Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, *71*(5), 1205–1222.

Guez, A. (2015). Sample-based Search Methods for Bayes-Adaptive Planning. *Unpublished PhD thesis.*

Gureckis, T. M., & Markant, D. (2009). Active learning strategies in a spatial concept learning game. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 3145–3150). Austin, TX: Cognitive Science Society.

Gureckis, T. M., & Markant, D. B. (2012, September). Self-Directed Learning: A Cognitive and Computational Perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., . . . Chan, P. (2016). psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behavior Research Methods*, *48*(3), 829–842.

Gutmann, M. U., Corander, J., et al. (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning*

Research.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models.* Wiley Online Library.

Hebb, D. O. (1955). Drives and the cns (conceptual nervous system). *Psychological Review*, *62*(4), 243.

Hoch, J. E., Rachwani, J., & Adolph, K. E. (in revision). Why infants move: Locomotor exploration is not directed toward destinations. *Child Development*.

Hofer, M., & Nelson, J. (2016). *Entropy-based approaches to Mastermind: Exploring human intuitions about Information and Uncertainty.* Poster presented at the Middle European Interdisciplinary Conference in Cognitive Science.

Huttenlocher, J. (1962). Some effects of negative instances on the formation of simple concepts. *Psychological Reports*, *11*(1), 35–42.

Kaiser, M. K., McCloskey, M., & Proffitt, D. R. (1986). Development of intuitive theories of motion: Curvilinear motion in the absence of external forces. *Developmental Psychology*, *22*(1), 67.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, *34*(7), 1185–243.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304.

Klenske, E. D., & Hennig, P. (2016). Dual control for approximate bayesian reinforcement learning. *Journal of Machine Learning Research*, *17*(127), 1–30.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Kretch, K. S., & Adolph, K. E. (2017). The organization of exploratory behaviors in infant locomotor planning. *Developmental science*, *20*(4).

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology.* Sage.

Lagnado, D. A., & Sloman, S. A. (2002). Learning causal structure. In *Proceedings of the 24^{th} Annual Meeting of the Cognitive Science Society.* Erlbaum.

Lagnado, D. A., & Sloman, S. A. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *30*, 856–876.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*(3), 451–60.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567.

Lucas, C. G., Bridgers, S., Griffiths, T. L., & Gopnik, A. (2014). When children are better (or at least more open-minded) learners than adults: Developmental differences in learning the forms of causal relationships. *Cognition*, *131*(2), 284–299.

MacKay, D. J. (1992). Information-based objective functions for active data selection. *Neural computation*, *4*(4), 590–604.

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, *24*(12), 2351–2360.

Markant, D., Dubrow, S., Davachi, L., & Gureckis, T. (2014). Deconstructing the effect of self-directed learning on episodic memory. *Memory & Cognition*, *42*(8), 1211-1224.

Markant, D. B., & Gureckis, T. (2014b). A preference for the unpredictable over the informative during self-directed learning. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Markant, D. B., & Gureckis, T. M. (2012). Does the utility of information influence sampling behavior? In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Markant, D. B., & Gureckis, T. M. (2014a). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

Marr, D. (1982). *Vision.* New York: Freeman & Co.

McCloskey, M. (1983). Naive theories of motion. *Mental models*, 299–324.

McCormack, T., Bramley, N. R., Frosch, C., Patrick, F., & Lagnado, D. A. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, *141*, 1–22.

Michotte, A. (1946/1963). *The perception of causality.* Basic Books.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., & Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition*, *130*(1), 74–80.

Oakes, L. M., & Cohen, L. B. (1990). Infant perception of a causal event. *Cognitive Development*, *5*(2), 193–207.

Oaksford, M., & Chater, N. (1994). A rational analysis ofthe selection task as optimal data selection. *Psychological Review*, *101*(4), 608–631.

Osman, M. (2011). *Controlling uncertainty: Decision making and learning in complex worlds.* Wiley.

Pearl, J. (2000). *Causality.* New York: Cambridge University Press (2nd edition).

Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, *497*(7447), 74–79.

Piaget, J. (1936). *Origin of intelligence in the child.* Penguin books.

Raiffa, H. (1974). *Applied statistical decision theory.* Wiley.

Ratmann, O., Jørgensen, O., Hinkley, T., Stumpf, M., Richardson, S., & Wiuf, C. (2007). Using likelihood-free inference to compare evolutionary dynamics of the protein networks of h. pylori and p. falciparum. *PLoS Computational Biology*, *3*(11), e230.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development*, *44*(2), 43–58.

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, *120*(2), 411.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1), 1–46.

Schulz, E., Klenske, E., Bramley, N., & Speekenbrink, M. (2017). Strategic exploration in human adaptive control. *bioRxiv*, 110486.

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: preschoolers engage in more exploratory play when evidence is confounded. *Developmental psychology*, *43*(4), 1045.

Shannon, C. E. (1951). Prediction and entropy of printed english. *The Bell System Technical Journal*, *30*, 50–64.

Siegel, M. H., Magid, R., Tenenbaum, J. B., & Schulz, L. E. (2014). Black boxes: Hypothesis testing via indirect perceptual evidence. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Smith, K. A., de Peres, F., Vul, E., & Tenenbaum, J. B. (2017). Thinking inside the box: Motion prediction in contained spaces uses simulation. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Smith, K. A., & Vul, E. (2014). Looking forwards and backwards: Similarities and differences in prediction and retrodiction. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1467–1472). Austin, TX: Cognitive Science Society.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task.

Cognition, *57*(1), 31–95.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search.* Cambridge, MA: MIT Press (2nd edition, 2000).

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infantsâĂŹ learning and exploration. *Science*, *348*(6230), 91–94.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453–489.

Tenenbaum, J. B., Griffiths, T. L., & Niyogi, S. (2007). Intuitive theories as grammars for causal inference. *Causal learning: Psychology, philosophy, and computation*, 301–322.

Tenenbaum, J. M., & Witkin, A. (1983). On the role of structure in vision. *Human and Machine Vision*, 481–543.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15.

Ullman, T., Stuhlmüller, A., Goodman, N., & Tenenbaum, J. B. (2014). Learning physics from dynamical scenes. In *Proceedings of the 36th Annual Meeting of the Cognitive Science Society.* Austin, TX: Cognitive Science Society.

Ullman, T., Stuhlmüller, A., Goodman, N., & Tenenbaum, J. B. (to appear). Learning physical parameters from dynamic scenes. *Cognitive Psychology*.

Ullman, T. D., Goodman, N. D., & Tenenbaum, J. B. (2012). Theory learning as stochastic search in the language of thought. *Cognitive Development*, *27*(4), 455–480.

Ullman, T. D., Spelke, E., Battaglia, P., & Tenenbaum, J. B. (2017, sep). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, *21*(9), 649–665. doi: 10.1016/j.tics.2017.05.012

Vul, E., Frank, M. C., Tenenbaum, J. B., & Alvarez, G. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In *Advances in neural information processing systems* (pp. 1955–1963).

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly journal of experimental psychology*, *20*(3), 273–281.

Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu: Annotationes anatomicae et physiologicae.* Koehler.

White, P. A. (1995). *The understanding of causation and the production of action: From infancy to adulthood.* Psychology Press.

Winer, B. J. (1962). Statistical principles in experimental design.

Wu, C. M., Schulz, E., Speekenbrink, M., Nelson, J. D., & Meder, B. (2017). Exploration and generalization in vast spaces. *bioRxiv*. doi: 10.1101/171371

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences*, *10*(7), 301–308.

Appendix

Physics simulator settings

We used a standard open source 2D physics simulator called Box2D (`http://box2d.org/about/`). Source code is available at `https://github.com/erincatto/Box2D`. The simulator is written in C but to integrate it with our Psiturk interface, we used a javascript port (`box2d-js.sourceforge.net`) of a Flash port (`http://box2dflash.sourceforge.net/`) of the original engine.

Demo code for our Experiments is available at `https://github.com/neilbramley/active_physics`. Table A1 details the settings for the physics simulator common to both experiments.[23,24]

Data

The response data from both experiments and code for analyses is available at `https://github.com/neilbramley/active_physics`.

Online information measures

We defined PD_{mass} as the average predicted divergence between worlds $w \in \mathcal{W}$ differing on the target mass dimension. To write this we split $\mathcal{W}$ into three subsets $\mathcal{W} = \mathcal{W}_A \cup \mathcal{W}_B \cup \mathcal{W}_{\text{same}}$ such that models $\mathcal{W}_A\{i\}, \mathcal{W}_B\{i\}$ and $\mathcal{W}_{\text{same}}\{i\}$ are identical on all dimensions except the target mass. We then evaluated their expected divergence by averaging over all comparisons (e.g. A vs. B, A vs. *same* and B vs. *same*) and all other properties (e.g. $i \in \frac{|\mathcal{W}|}{3}$), using the same Gaussian error assumption as Ullman et al. (2014, to appear) and Equation 1. To get a measure that increases for greater average divergences (unlike the likelihoods that decreased), we subtracted these scores from 1. The resulting average divergence can be written as:

[23] Damping in Box2D slows objects while they are not in contact with any other objects (like wind resistance). The controlled object was given high damping to prevent it from oscillating for a long time around the cursor location.

[24] Friction in Box2D occurs when two objects slide past each other while touching (e.g. a puck sliding along a boundary wall).

Table A1
Physics world settings

Parameter	Value
N frames	2700 (Exp 1), 1800 (Exp 2)
Trial length	45s (Exp 1), 30s (Exp 2)
Box2D step size	1/60s ($\approx 17ms$)
Pixels to meters	100
Object velocity cap	30 $^{\mathrm{m}}/_{\mathrm{s}}$
Refresh criterion	Fastest object < 0.25 $^{\mathrm{m}}/_{\mathrm{s}}$
Pause on refresh	500ms
Starting velocities	(x, y) drawn from Unif$(-10, 10)$ $^{\mathrm{m}}/_{\mathrm{s}}$
World width	6m (600 pixels)
World height	4m (400 pixels)
Attractive forces	$+3$ $^{\mathrm{m}}/_{\mathrm{s}^2}$
Repulsive forces	-3 $^{\mathrm{m}}/_{\mathrm{s}^2}$
Controlled object attraction cursor	$.2 \times \mathrm{dist}(\mathrm{cursor}, \mathrm{object})$ $^{\mathrm{m}}/_{\mathrm{s}^2}$
Controlled object damping	10
Puck masses	1kg (2kg for heavy target)
Puck friction	.05
Puck elasticity	.98
Puck damping	.05
Puck radius	0.25 m
Puck object types	Dynamic
Wall mass	n/a
Wall friction	.05
Wall elasticity	.98
Wall damping	n/a
Wall width	0.2m
Wall object types	Static

$$\underset{\text{mass}}{\mathrm{PD}} = 1 - \underset{\mathcal{I}<\mathcal{J}\in[\mathcal{W}_A, \mathcal{W}_B, \mathcal{W}_{\text{same}}]}{\mathbb{E}} \left[\underset{i\in\mathcal{I}, j\in\mathcal{J}}{\mathbb{E}} \left[e^{-\frac{\eta}{2}(\mathbf{s}^t - \mathbf{d}^t)^\top \Sigma^{-1} (\mathbf{s}^t - \mathbf{d}^t)} \right] \right]. \tag{A-1}$$

We do the same for $\mathrm{PD}_{\text{force}}$, replacing $\mathcal{W}_A, \mathcal{W}_B$ and $\mathcal{W}_{\text{same}}$ with $\mathcal{W}_{\text{attract}}, \mathcal{W}_{\text{repel}}$ and $\mathcal{W}_{\text{none}}$:

$$\underset{\text{force}}{\mathrm{PD}} = 1 - \underset{\mathcal{I}<\mathcal{J}\in[\mathcal{W}_{\text{attract}}, \mathcal{W}_{\text{repel}}, \mathcal{W}_{\text{none}}]}{\mathbb{E}} \left[\underset{i\in\mathcal{I}, j\in\mathcal{J}}{\mathbb{E}} \left[e^{-\frac{\eta}{2}(\mathbf{s}^t - \mathbf{d}^t)^\top \Sigma^{-1} (\mathbf{s}^t - \mathbf{d}^t)} \right] \right]. \tag{A-2}$$

Finally, to compute baseline PD we repeat this procedure for all 7 dimensions of the problem $\forall$ z $\in Z$ (e.g. the target mass, target force and the five possible distractor forces) and take the average of all of these:

$$\underset{\text{baseline}}{\text{PD}} = 1 - \underset{z\in Z}{\mathbb{E}}\left[\underset{\mathcal{I}<\mathcal{J}\in[\mathcal{W}_{z_1},\mathcal{W}_{z_2},\mathcal{W}_{z_3}]}{\mathbb{E}}\left[\underset{i\in\mathcal{I},j\in\mathcal{J}}{\mathbb{E}}\left[e^{-\frac{\eta}{2}(\mathbf{s}^t-\mathbf{d}^t)^\top\Sigma^{-1}(\mathbf{s}^t-\mathbf{d}^t)}\right]\right]\right]. \quad \text{(A-3)}$$

We assumed a different scaling parameter $\eta = 10$ for the predictive divergence measures (Equations A-1, A-2 and A-3) rather than the $\beta = \frac{1}{50}$ for computing model posteriors (Equation 1). Using the same parameter for both led to underflow in the case of PD because, while the overall likelihood is a product over all frames, the predicted divergences are computed per frame then averaged for each ten frame snap-back window. In other respects, a range of values for β and η did not affect the reported comparisons.

Free response coding

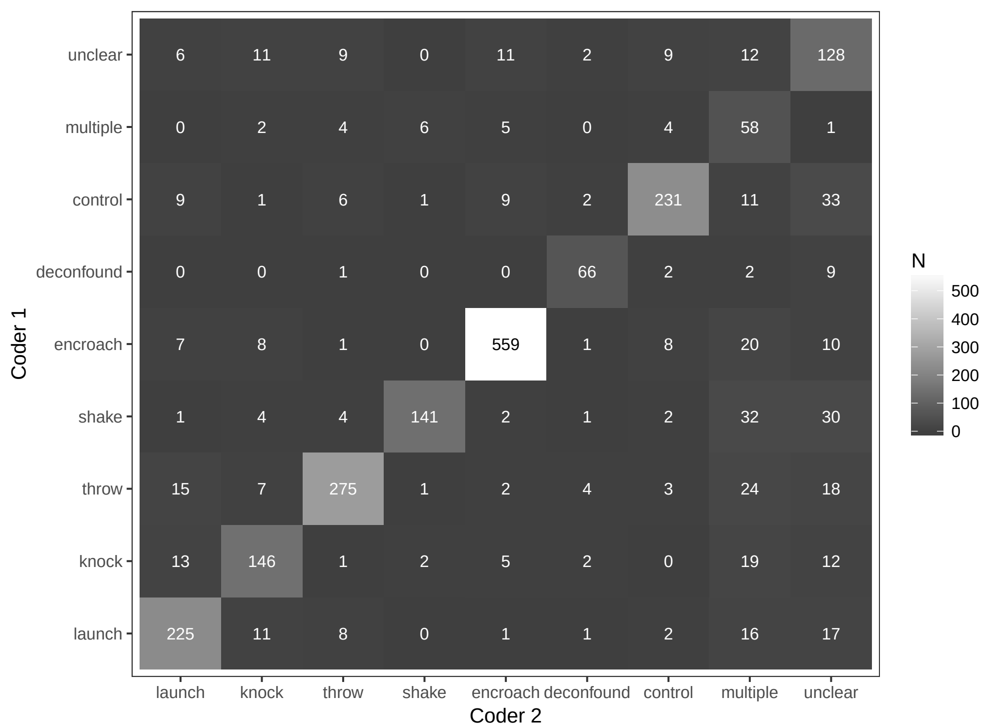


Figure A1. Confusion matrix for our two independent strategy coders, showing close correspondence between coders and no clear systematicity to disagreements with the exception of a greater tendency for coder 2 to assign multiple or unclear.

Comparing different distance measures for likelihood approximation

Figures A2 and A3 illustrate the differences in information model predictions depending which aspects of physical dynamics are used to calculate likelihoods and predictive divergences. Figures A2 compares measures at a high level of aggregation showing that location information x, y largely yields results similar to those from motion information r, θ but results in somewhat lower posterior uncertainty about masses. We include $\log r$ because $\Delta \log r$ encodes relative magnitude information in line with Weber's law (1834). That is, a doubling of magnitude has the same $\Delta \log r$ regardless of the absolute values involved. However, for these data $\log r$ proved relatively non-diagnostic about both key properties. Focusing only on θ also results in high uncertainty about both properties and as such, both of these measures might resolve the inconsistency between human and model accuracy about masses. However they also fail to capture pattern shown by the other measures and combinations in Experiment 2 A2b, whereby participants generate more information about mass when this matches their goal.

Figure A3 compares the measures at a low level, in terms of the instantaneous predictive divergence they predict across the trial used in Figure 7 in the main text. We include plots showing these variants for all participants and trials paired with the replay videos for Experiment 2 in the online supplement available at `https://github.com/neilbramley/active_physics`.

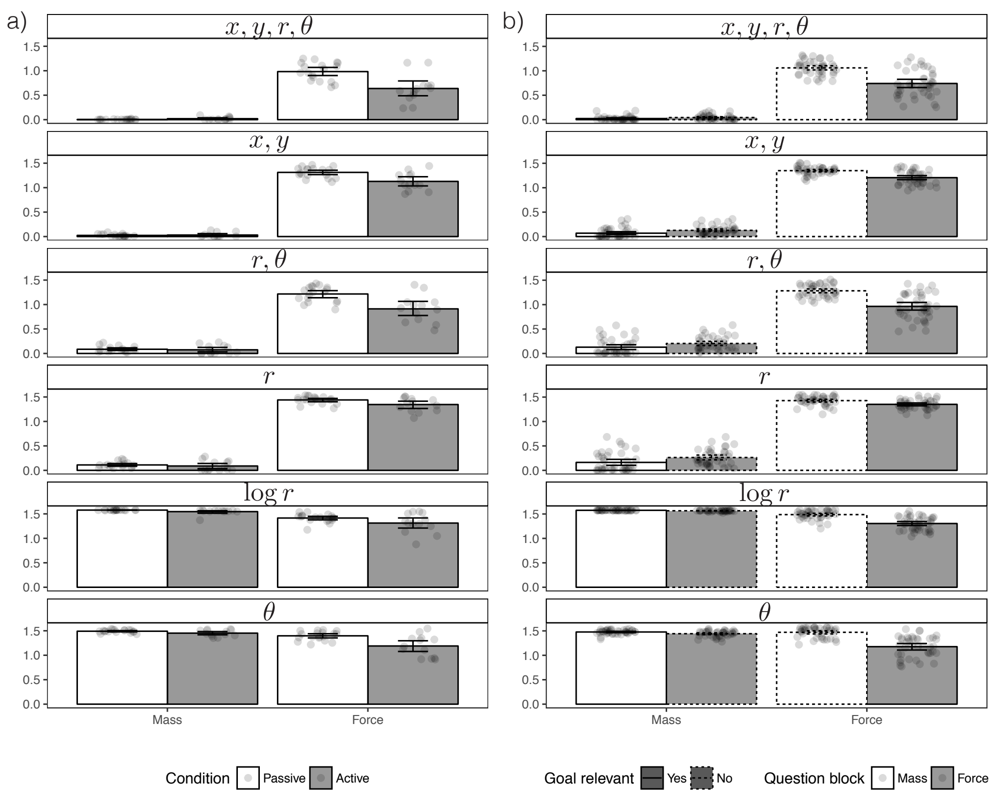


Figure A2. Comparison of entropies resulting from combining different distance measures with Equation 1. a) Experiment 1: Comparing mass and force entropy across passive and active trials as in Figure 5. b) Experiment 2: Comparing mass and force entropy across mass and force question blocks as in Figure 10.

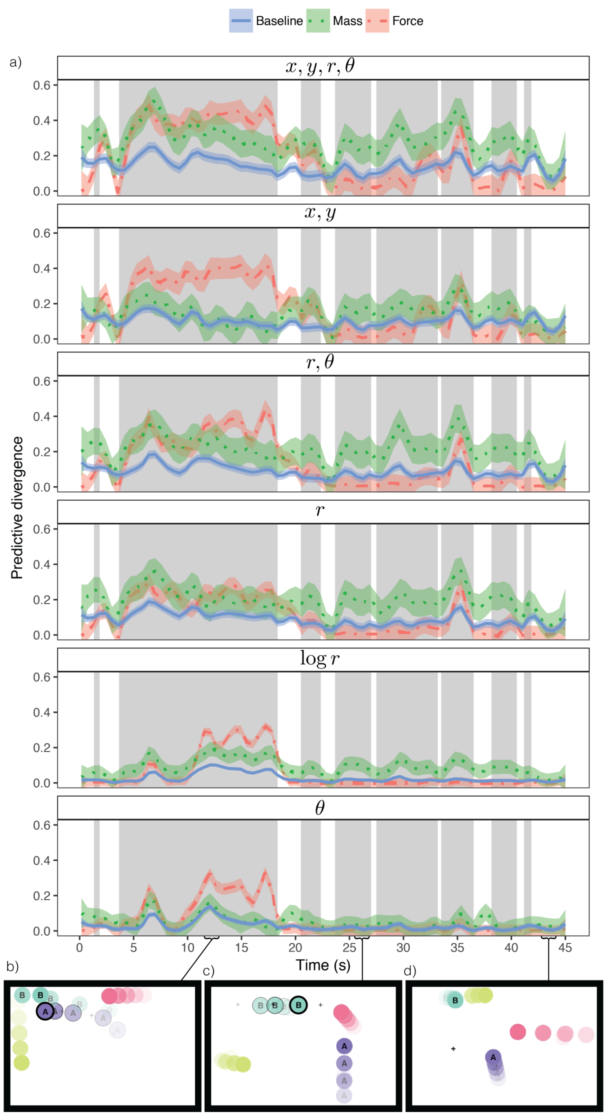


Figure A3. Example trial from Experiment 1 as in Figure 7. Replay: `https://neilrbramley.com/experiments/apl/e1/replays?p=1` a) Timelines for PD using various combinations of distance measure. b–d) As in Figure 7, visualizations of actions during clip. A sequence of screen-shots are superimposed with later frames becoming more opaque. Thick black circles indicate controlled object and "+" indicates the mouse.